

https://dspace.org/

Chronic non-communicable disease multimorbidity in South African adults: evidence from the who sage study

Chidumwa, Glory

2022

https://hdl.handle.net/10539/34340

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

CHRONIC NON-COMMUNICABLE DISEASE MULTIMORBIDITY IN SOUTH AFRICAN ADULTS: EVIDENCE FROM THE WHO SAGE STUDY

by

GLORY CHIDUMWA (Student number: 1332997)

Supervisors:

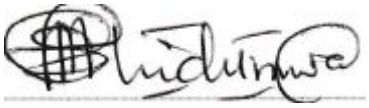
Dr Lisa J Ware, Prof Lisa Micklesfield, and Dr Innocent Maposa

Original published work submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy (Public Health in the field of Biostatistics)

April 2022 (Johannesburg)

PLAGIARISM DECLARATION

I, Glory Chidumwa, declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy (Public Health in the field of Biostatistics) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Glory Chidumwa

April 2022

DEDICATION

To the **Chidumwa** family, my love - **Mutsa Pretty Makamanzi** and my friends: For your advice, unwavering support, belief in me, love and patience, the least I can say is thank you!

Scientific outputs from this thesis

Two published manuscripts and one in draft are included as part of the thesis:

1. **Published:** Chidumwa G, Maposa I, Corso B, Minicuci N, Kowal P, Micklesfield LK, Ware LJ (2021) Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2. BMJ Open 11 (1):e041604
2. **Published:** Chidumwa G, Maposa I, Kowal P, Micklesfield LK, Ware LJ (2021) Bivariate Joint Spatial Modelling to Identify Shared Risk Patterns of Hypertension and Diabetes in South Africa: Evidence from WHO SAGE South Africa Wave 2. International Journal of Environmental Research and Public Health 18 (1):359
3. **Draft manuscript:** Chidumwa G, Maposa I, Kowal P, Micklesfield LK, Ware LJ Understanding the inter-relationships between socio-economic, socio-demographic, behavioural, and environmental factors for multimorbidity: A structural equation modelling approach.

Conferences attended

1. Sub-Saharan African Network (SUSAN) of the International Biometrics Society (IBS) and DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB), **8th - 11th of September 2019**, Capetown, South Africa
2. DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) - *virtual*, **20 October 2020**

Academic trip (Data exploration and Initial analysis)

Neuroscience Institute, National Research Council, Padua, Italy**. 28 October – 7 November 2019**

THESIS ABSTRACT

Background

Non-communicable diseases (NCDs) are the leading cause of global mortality, and in South Africa were estimated to account for 57.4% of the total burden of disease in 2016. Within an individual, the co-existence of two or more chronic (at least three months) non-communicable, mental health or infectious disease is referred to as 'multimorbidity' (MM). While risk factors for MM are present across the life course, the onset of most NCDs occurs in middle to older age. However, there is limited research that explores MM among middle aged and older adults in South Africa.

Aim

The aim of this thesis was to investigate MM in middle-aged and older South African adults from the WHO SAGE cohort. This was addressed in three parts:

1. To determine the prevalence of multimorbidity and the co-occurrence of chronic diseases in a cohort of South African adults over the age of 50 years, and to identify the demographic, anthropometric and behavioural factors associated with different multimorbidity clustering.
2. To examine the spatial distribution of hypertension and diabetes jointly, and the distribution of shared unmeasured characteristics on hypertension and diabetes in South African middle aged and older adults.
3. To determine the complex inter-relationships between socio-economic, socio-demographic, behavioural, and environmental factors associated with multimorbidity and depression in South African middle aged and older adults.

This study will contribute to an expansion of the epidemiology of NCDs and biostatistical literature through the novel application of statistical techniques such as latent class analysis (LCA), bivariate joint shared component modelling, and the generalized structural equation model (gSEM) approach, used to address the research objectives above.

Methods

Cross-sectional secondary analysis of data collected as part of a panel study carried out by the WHO SAGE Wave 2 in South Africa in 2015 was completed. The current thesis included adults (≥18 years old for objective 1 and ≥40 years old for objectives 2 and 3) for whom data on 7 NCDs (angina, arthritis, asthma, chronic lung disease, depression, diabetes, and hypertension) and socioeconomic, demographic, behavioural, and anthropometric information were available. Further details of the South African SAGE sample are given in separate methods sections. Latent class analysis was used to identify groups and determine the co-occurrence of the NCDs.

Bivariate joint shared component modelling was used to assess the clustering and association between diabetes and hypertension and to jointly model the shared and disease-specific geographical variation of hypertension and diabetes. Lastly, I utilized the logit models and gSEM to explore the association between socioeconomic, demographic, and behavioural factors, and multimorbidity and depression.

Results

The study used the WHO SAGE South Africa Wave 2 data collected in 2015 on 2761 participants aged 18 years and above. The majority of the sample were female (n=1846; 67%) The prevalence of multimorbidity was 21%. The LCA identified three latent classes which were named as follows: *minimal MM risk* (83%), *concordant MM (i.e., expected, or typical clustering of hypertension and diabetes;* 11%), and *discordant MM (less typical clustering of combination of angina, asthma, chronic lung disease, arthritis and depression;* 6%). Using the *minimal MM risk* group as the reference, female [Relative risk ratio (RRR) = 4.57; 95% Confidence Interval (CI) (1.64; 12.75); p-value=0.004] and older [RRR=1.08; 95% CI (1.04; 1.12); p-value<0.001] participants were more likely to belong to the *concordant MM* group. Tobacco users [RRR= 8.41; 95% CI (1.93; 36.69); p-value=0.005] and older [RRR=1.09; 95% CI (1.03; 1.15); p-value=0.002] participants had a higher likelihood of belonging to the *discordant MM* group.

As hypertension and diabetes commonly co-occur in South African adults the second study modelled the shared and disease-specific spatial distribution of these two NCDs using bivariate joint shared component modelling. The shared component of diabetes and hypertension had distinct spatial patterns with higher odds in the eastern districts of Kwa-Zulu Natal and central Gauteng province of South Africa. The shared component represents unmeasured influences such as health behaviour characteristics or social determinants of health in our population. My study further showed that the shared component for hypertension and diabetes, which may include ecological factors and environmental determinants such as population density, pollution, transport, power, and local food environment is more pronounced in certain South African provinces such as Gauteng and Kwa-Zulu Natal.

Using logistic regression and generalized structural equation modelling (gSEM) to explore the associations between socio-economic, socio-demographic, behavioural and environmental factors, and risk of depression and multimorbidity, the results were as follows: In the unadjusted logistic regression analyses, feeling “unsafe” [aOR=2.04; 95% Confidence Interval: 1.25; 3.42], being female, [aOR=1.93; 95% Confidence Interval: 1.02; 3.62], and older age [aOR=1.05; 95% Confidence Interval: 1.02; 1.08] were associated with higher odds for multimorbidity. In addition, being female, belonging to the highest wealth tertile relative to those in the lowest

tertile, and living in an urban area were significantly associated with higher odds of depression [OR=1.39; 95% Confidence Interval: 0.59; 3.29]. Similarly, in the gSEM model, where models are estimated concurrently, demographic factors [older age (aOR=1.03, 95% Confidence Interval: 1.01; 1.05) and being female (aOR= 3.02; 95% Confidence Interval: 1.88; 4.86)] and behavioural factors [individuals with history of tobacco avoidance (aOR=0.46; 95% Confidence Interval: 0.27;0.75), and good sleep quality (aOR=0.59; 95% Confidence Interval: 0.39;0.91)] were significantly associated with multimorbidity. Moreover, using the gSEM approach, multimorbidity had two-fold odds of depression and was statistically significant (aOR=2.41; 95% Confidence Interval: 1.36;4.28).

Discussion and conclusion

Results indicate that in my sample of middle aged and older South African adults 1 in 5 people above the age of 45 years have MM. Risk factors for multimorbidity included older age, female sex and tobacco use, but results show that these may differ depending on whether the diseases are concordant or discordant, which may suggest different avenues for intervention. Given the co-occurrence of NCDs, I underscore the need for the healthcare system to focus on managing multiple diseases rather than a vertical approach in managing single diseases, specifically for hypertension and diabetes. In addition, policy-makers may potentially use our spatial results for purposes of more localised resource allocation and prevention health programs in high burden hypertension and diabetes areas in South Africa. In addition, future efforts should focus on understanding the unmeasured shared component, which may include infectious diseases, or frequently co-occurring common conditions, and to evaluate clustering patterns.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to GOD without whom this study would not have been completed.

Immeasurable gratitude and sincere appreciation for help and support are extended to the following:

Dr Lisa J Ware, Prof Lisa Micklesfield, and **Dr Innocent Maposa** for their guidance and encouragement in completing the thesis.

I have had financial support from the **Developing Excellence in Leadership, Training and Science (DELTAS)** Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences' (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (grant 107754/Z/15/Z- DELTAS Africa **Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB)** programme) and the UK government. The WHO-SAGE multicountry study is supported by the WHO and the Division of Behavioral and Social Research (BSR) at the National Institute on Aging (NIA), US National Institutes of Health, through Interagency Agreements (OGHA04034785; YA1323-08-CN-0020; Y1-AG-1005-01) with WHO, a Research Project Grant R01AG034479, and in-kind support from the South Africa Department of Health.

World Health Organization SAGE team and **Neuroscience Institute, National Research Council, Italy team (Dr Nadia Minicuci, Dr Barbara Corso and Illaria Rocco)** for their support in data access and exploration.

My family, lecturers, friends from the department and church for their worthy support and co-operation always.

TABLE OF CONTENTS

PLAGIARISM DECLARATION ii
DEDICATION iii
THESIS ABSTRACT v
ACKNOWLEDGEMENTS viii
LIST OF ABBREVIATIONS xi
GLOSSARY OF TERMS xii
LIST OF TABLES xiii
LIST OF FIGURES xiv
CHAPTER 1: LITERATURE REVIEW 1
1.1 The Burden of NCDs in South Africa and globally 1
1.2 Multimorbidity 2
1.3 Prevalence of MM 2
1.4 Factors associated with multimorbidity. 5
1.5 Novel application of statistical methods 7
1.6 Problem statement 9
1.7 Justification 9
1.8 Aims and objectives 10
CHAPTER 2: METHODS 12
2.1 Statistical methods 13
2.1.1 Latent class analysis 13
2.1.2 Bivariate joint spatial modelling 16
2.1.3 Structural equation modelling 18
CHAPTER 3: IDENTIFYING CO-OCCURRENCE AND CLUSTERING OF CHRONIC DISEASES USING LATENT CLASS ANALYSIS: CROSS-SECTIONAL FINDINGS FROM SAGE SOUTH AFRICA WAVE 2 20
Authors' Contributions: 21
3.1 Introduction 21
3.2 Methods 23
3.3 Results 25
3.4 Discussion 30
CHAPTER 4: BIVARIATE JOINT SPATIAL MODELLING TO IDENTIFY SHARED RISK PATTERNS OF HYPERTENSION AND DIABETES IN SOUTH AFRICA: EVIDENCE FROM WHO SAGE SOUTH AFRICA WAVE 2 33
Author Contributions 33
4.1 Introduction 34
4.2 Materials and Methods 35
4.3 Results 39
4.4 Discussion 45
4.5 Conclusion 47
CHAPTER 5: UNDERSTANDING THE INTER-RELATIONSHIPS BETWEEN SOCIO-ECONOMIC, SOCIO-DEMOGRAPHIC, BEHAVIOURAL, AND ENVIRONMENTAL FACTORS FOR MULTIMORBIDITY: A STRUCTURAL EQUATION MODELLING

APPROACH **48**
Author Contributions 48
5.1 Introduction 49
5.2 Methods 50
5.3 Results 53
5.4 Discussion 56
CHAPTER 6: DISCUSSION AND CONCLUSIONS **60**
6.1 Overview of research question and aim of the thesis 60
6.2 Summary of objectives and key findings 61
6.3 Summary and implications of study findings 62
6.4 Limitations and strengths 64
6.5 Future research and recommendations 65
6.6 Conclusions 67
REFERENCES **101**
APPENDICES **115**
Appendix A: Summary of NCD's missing data. **115**
Appendix B: Syntax for Paper 1. 116
SAS code 116
Stata code 118
Appendix C: Syntax for Paper 2 128
OpenBugs code 128
Stata code 133
Appendix D: Stata code for Paper 3 136
Appendix E: Ethical clearance certificate 140
Appendix F: Supplementary tables and figures 141
Appendix G: Declaration: Student's contribution to articles and agreement of co-authors 143
Appendix H: WHO-SAGE generic questionnaire 145

LIST OF ABBREVIATIONS

WHO - World Health Organization

SAGE - Study on global AGEing and adult health

NCDs - non-communicable diseases

LCA – latent class analysis

gSEM – generalized structural equation modelling

LMICs - low- and middle-income countries

BP – blood pressure

NIDS - The National Income Dynamics Study

HIV - human immunodeficiency virus

TB - tuberculosis

CVD - cardiovascular disease

COPD - chronic obstructive pulmonary disease

ICDM - integrated chronic disease management

MM - multimorbidity

UN – United nations

GLM - generalized linear model

DBP – diastolic blood pressure

SBP - systolic blood pressure

SASH - The South African Stress and Health

NICE - National Institute for Health and Care Excellence

GHS - General Household Surveys

UN – United Nations

GLOSSARY OF TERMS

Multimorbidity – the co-existence of two or more chronic non-communicable, mental health or infectious diseases of long duration, at least three months, in the same person.

Non-communicable diseases - a disease that is not transmissible directly from one person to another.

Latent class analysis - a statistical method for identifying unmeasured class membership among subjects using categorical and/or continuous observed variables.

Generalized Structural Equation Modelling - a family of statistical techniques utilized in the analysis of multivariate, categorical, and ordinal data in order to measure latent variables and their connection with each other.

Shared component model – a multivariate model for joint spatial analysis that separates the underlying risk surface for each disease into a shared component - a surrogate for unobserved covariates that display spatial structure and are common to both diseases, and a disease-specific component.

LIST OF TABLES

Table 3. 1: Characteristics of participants by latent class category (n = 1967).....28
Table 3. 2: Results from multinomial logistic regression for factors associated with latent class membership.....29

Table 4. 1: Prevalence of hypertension, diabetes, and comorbidity (having both hypertension and diabetes or not), by province (N = 2761). 39
Table 4. 2: Summary statistics for demographic, socioeconomic status, anthropometry, behavioral and blood pressure characteristics, by comorbidity, N = 2761. 40
Table 4. 3: Results from multivariate logistic regression model for hypertension and diabetes by demographic, socioeconomic, anthropometric, and behavioral factors. 42

Table 5. 1: Prevalence of multimorbidity presented by explanatory factors.....54
Table 5. 2: Results of fitting multivariable logistic regression models for factors associated with depression and multimorbidity.....55

Supplementary Table 3. 1: Latent class analysis fit statistics 141

Supplementary Table 4. 1: Population size and sample size, by province.....142
Supplementary Table 4. 2: Results of fitting stepwise backward elimination multiple logistic regression models for factors associated with hypertension and diabetes (p=0.10).....142

LIST OF FIGURES

Figure 1. 1: The epidemiological change in South Africa between 1997 and 2016.....2
Figure 1. 2: The NCD causation pathway.....6

Figure 3. 1: Study flow diagram.....26
Figure 3. 2: Prevalence of chronic NCDs by sex.....27

Figure 4. 1: The South African map.....36
Figure 4. 2: Odds of disease-specific components for (a) hypertension, and (b) diabetes, by district, World Health Organization (WHO) Study on Global Ageing and Adult Health (SAGE) South Africa Wave 2. 43
Figure 4. 3 (a): Odds of the shared component in the joint (hypertension and diabetes) shared model, (b,c): Odds of the shared component risk contribution maps for hypertension and diabetes, by district, WHO SAGE South Africa Wave 2. 44

Figure 5. 1: Conceptual framework for the study 52
Figure 5. 2: Generalized structural equation model results for the hypothesized model. 56

Supplementary Figure 3. 1: Prevalence of NCDs, by latent class.....141

CHAPTER 1: LITERATURE REVIEW

The first chapter of the thesis gives a literature review of the NCDs burden globally and in South Africa, definition of multimorbidity (MM), prevalence of MM, factors associated with MM, novel application of statistical methods, problem statement and justification, and aim and specific objectives of the thesis.

1.1 The Burden of NCDs in South Africa and globally

The NCD burden is not only high but has become the leading cause of death globally [1]. NCDs also have contributed to over two thirds of the global mortality rate [2] and are the leading cause of global disability-adjusted life-years [3]. In 2015 NCDs resulted in 71% of deaths globally [4]. The existing high burden of NCDs is associated with premature death, loss of quality of life and has negative economic impacts on families and society as a whole [5]. As shown in the Scottish study by Barnett et al [6], the onset of multimorbidity in developed countries may occur 10 to 15 years earlier in individuals living in deprived areas [7,8]. Additionally, NCD deaths are projected to continue to rise worldwide, with the greatest increase expected to be seen in LMICs [2] where NCDs include hypertension, diabetes, depression, arthritis, asthma, angina or angina pectoris, chronic lung disease, stroke, and vision impairment [9]. In South Africa, 57.4% of the total burden of disease in 2016 was attributed to NCDs [10,11].

In comparison, heart disease, diabetes and stroke together constitute the second most important cause of death in adult South Africans and NCDs accounted for 39% of deaths and for considerable premature mortality in South Africa in 2010 [12,13]. Figure 1.1 shows the epidemiological change in South Africa between 1997 and 2016. As shown, South Africa is in the midst of a profound health transition that is characterised by a quadruple burden of communicable, non-communicable, perinatal and maternal, and injury-related disorders in addition to the high prevalence of communicable diseases such as human immunodeficiency virus (HIV), tuberculosis (TB), and COVID-19 [14]. Non-communicable diseases are emerging in both rural and urban areas, most prominently in poor people living in urban settings, and are resulting in increasing pressure on acute and chronic health-care services [13].

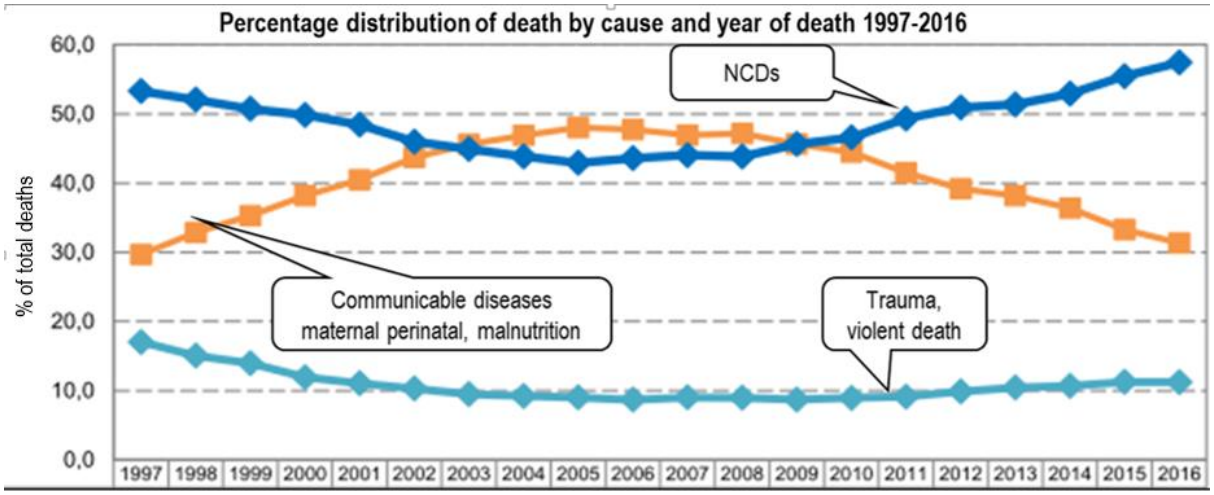


Figure 1. 1: The epidemiological change in South Africa between 1997 and 2016

Source: Statistics South Africa (Available at: https://www.sancda.org.za/ncds-increasingly-the-main-cause-of-death-in-south-africa-time-for-action/)

1.2 Multimorbidity

The definition of multimorbidity in research differs from one study to another. In the current study, MM is defined as "the co-existence of two or more chronic conditions". According to the Academy of Medical Sciences, the term 'multiple long-term conditions' is also used to refer to the existence of two or more long-term conditions in a single individual [15]. One of which is either: a physical non-communicable disease of long duration, such as cardiovascular disease or cancer or a mental health condition of long duration, such as a mood disorder or dementia or an infectious disease of long duration, such as HIV or hepatitis C [15]. The UK National Institute for Health and Care Excellence (NICE) uses a slightly expanded version of this definition, and states that multimorbidity is the presence of two or more long-term health conditions, which can include: defined physical or mental health conditions such as diabetes or schizophrenia; ongoing conditions such as a learning disability; symptom complexes such as frailty; chronic pain or sensory impairment such as sight or hearing loss; or alcohol or substance misuse [15].

1.3 Prevalence of MM

Evidence suggests MM prevalence globally is increasing or MM prevalence reported is highly dependent on the included conditions. Folb and colleagues conducted a study to describe multimorbidity, related risk factors, disease severity and treatment status of patients with four important NCDs (hypertension, diabetes, chronic respiratory

disease and depression) attending public sector primary health care clinics in Eden and Overberg districts of the Western Cape in 2011 in SA. In this study, out of the total of 4 393 adults (18 years and older) enrolled from 38 clinics, the prevalence of MM was 42% [16]. In addition, Garin and colleagues analysed pooled data from two large multi-country studies: the Collaborative Research on Ageing in Europe project (Finland, Poland, and Spain) and the World Health Organization's Study on Global Ageing and Adult Health (China, Ghana, India, Mexico, Russia, and South Africa). The cross-sectional study obtained data from 41,909 noninstitutionalized adults older than 50 years and MM was defined as the presence of at least two NCDs namely: angina, asthma, chronic obstructive pulmonary disease, diabetes, obesity, hypertension, arthritis and depression. The prevalence's of NCDs MM in this study in South Africa, Ghana, India and China were 68%, 48%, 58% and 45% respectively [17]. The large variation in prevalence may be the result of differences between populations (dissimilar study populations or data sources, usually entailing differences in demographic characteristics and disease types or classification) or may also be due to variability in data available for different conditions and how they were measured [18-20].

The presence of multimorbidity, a condition where an individual has two or more co-existing conditions is becoming alarmingly more prevalent globally [21,22]. Though globally, the burden of multimorbidity has been focused on developed countries, there is increasing recognition that the burden of multimorbidity is rising in lower and middle-income countries [23]. This is largely due to the large reductions in mortality during childhood and childbirth, and reductions in deaths from infectious diseases in LMICs which has led to people living into their 60s and beyond [6,24,25]. Moreover, substantive successes in public health over the past several decades in these LMICs countries has led not only to marked gains in life expectancy but an increase in the prevalence of chronic conditions placing additional strain on healthcare systems that are ill-equipped for coping with the growing widespread demand for chronic disease care [26]. A scoping review by Eyowas and colleagues showed that the magnitude of multimorbidity in LMICs was reported to be between 14% and 68% and was projected to rise over the past recent years between 2015 and 2018 [27]. Wang and colleagues carried out a cross-sectional study to examine the prevalence and patterns of multimorbidity on adults (n=21 435) in north-eastern China. Multimorbidity was defined as having two or more of 18 specified prevalent chronic diseases. Approximately a quarter (24.7%) of the adults were found to be multimorbid for chronic diseases [28]. Another study was conducted

by Nunes and colleagues to evaluate the occurrence and factors associated with multimorbidity among Brazilians aged 50 years and over, and multimorbidity was assessed from a list of 19 morbidities. From the total of 9,412 individuals, 67.8% (95%CI 65.6–69.9) and 47.1% (95%CI 44.8–49.4) showed ≥ 2 and ≥ 3 diseases, respectively [29]. A hospital-based survey in an urban setting in Ghana interviewed adult patients (aged 18 years and above) attending a routine outpatient clinic at an inner-city hospital in Accra and reported a multimorbidity prevalence of 38.8%, with 48.6% of the patients with multimorbidity aged between 18–59 years old [30]. Highlighting the MM in South Africa, Afshar *et al* compared the prevalence of multimorbidity. MM was defined as the presence of two or more of the six conditions: arthritis, angina or angina pectoris (a heart disease), asthma, depression, schizophrenia or psychosis, and diabetes. Their study showed that the prevalence of MM in South Africa was 21.6% among 50–64-year-olds, considerably lower than the 30.1% prevalence observed in those 65 years and older [31]. Another SA study was carried out by Lalkhen, and colleagues aimed at evaluating the extent of multimorbidity among patients with NCDs in South African (SA) primary healthcare (PHC) [10]. A dataset obtained from a previous morbidity survey of SA ambulatory PHC was analysed including the following conditions: COPD, osteoarthritis, diabetes, asthma, hypertension, epilepsy, HIV, TB, depression, and anxiety disorders. Altogether 18 856 consultations were included in the dataset. Overall, 48.4% of patients had comorbidity and 14.4% multimorbidity. In an integrative literature review by Roomaney and colleagues to review prevalence studies of multimorbidity in South Africa, the prevalence of multimorbidity was low to moderate (3%–23%) in studies that included younger people (18 to 30 years) or had a wide range of selected age groups; and moderate to high (30%–87%) in studies of older adults (50 years and older) [32]. The multimorbidity prevalence in this thesis of 21% was dissimilar to others given that our study included younger participants (from 45 years of age). Another systematic review of studies in developed countries, indicated a prevalence ranging from 3.5% to 98.5% in primary care settings, and 13.1% to 71.8% [33]. In addition to distinct age groups being studied, this broad variation in prevalence may have occurred due to the fact that each study may assess a different set of chronic diseases and also how they were measured (self-report vs measured). Moreover, multimorbidity prevalence may vary from study to study due to differences in settings, populations and definitions of multimorbidity.

In their analysis, comorbidity was due to the presence of another disease at the consultation in addition to the selected NCD, while multimorbidity referred to the presence of two or more such diseases [10]. However, research on the prevalence, patterns and determinants of

multimorbidity, in South Africa is sparse. Systematic reviews on multimorbidity conducted to date were mostly on older adults and in high-income countries, reporting a pooled prevalence of multimorbidity between 38% and 66% [34-36]. Only two reviews included or were conducted on multimorbidity in LMICs [36,37].

In a country like South Africa, with a growing burden of NCDs and a high prevalence of HIV, information on multimorbidity can improve planning for healthcare delivery and utilisation, and reduce costs in the context of constrained health resources [38]. Non communicable diseases are responsible for 43% of deaths per year in South Africa, with most deaths being premature (deaths occurring before the age of 65 years) [39-41].

Despite such evidence on MM, traditionally NCDs are managed separately, without adequate consideration of comorbidity in individual patients. In a study by *Bosire et al* the patients interviewed often had to visit the health facilities multiple times to receive medication for each treatment, costing them time, effort and lost wages [42].

1.4 Factors associated with multimorbidity.

Age is a well-accepted risk factor for multimorbidity [43,44]. With increasing age numerous underlying physiological changes occur, and the risk of chronic disease rises with an increased risk of experiencing more than one chronic condition at the same time [45]. However, more recently studies have shown that multimorbidity is becoming more prevalent in middle-aged adults [25,46]. In addition, Taylor et al reported that in their study 70% of people with multimorbidity were less than 65 years of age [18].

In LMICs undergoing epidemiological transition, the prevalence of multimorbidity increases with greater affluence [23,46,47]. In contrast, in higher-income countries, the least economically advantaged are typically at greatest risk of multimorbidity [23,48]. This discrepancy may be because LMIC participants from lower socioeconomic background are less likely to report their health condition probably because of limited access to healthcare [49]. Data from the nationally representative annual South African General Household Surveys (GHS) between 2005 and 2008, reported that multimorbidity was associated with lower economic status [50]. In a study by Tayloe et al., multimorbidity was assessed across three age groups (20 to 39, 40 to 59, and 60 and above) and showed that participants aged 20-39 years with lower education levels were statistically significantly more likely to have multiple morbidities compared to those

with higher education level [18].

Individual risk factors linked to multimorbidity include obesity, inactivity, smoking and excessive use of alcohol; psychosocial factors, such as negative life events, stress, unemployment, external locus of control and small social network; mental health problems such as depression, and long-term treatments, such as being on antiretroviral therapy [23,46,51]. It has been reported that 40-59 year olds with multimorbidity are 1.71 times more likely to be current smokers than non-smokers [18]. Furthermore, among those with multimorbidity, a strong relationship exists between smoking status and the likelihood of depressive symptoms [18,52]. Obesity, has also previously been shown to be associated with multimorbidity [53,52] as obesity is closely associated with common NCDs such as type 2 diabetes [53,52]. Although less consistent, there is growing evidence on the sex differences in the prevalence of multimorbidity. Globally, multimorbidity in many studies has been reported to be more common in females than in males, depending on setting [52,53]. For this thesis, the NCD causation pathway conceptual framework was applied to data available from the World Health Organization Study on Global Ageing and Adult Health (SAGE) (Figure 1.2).

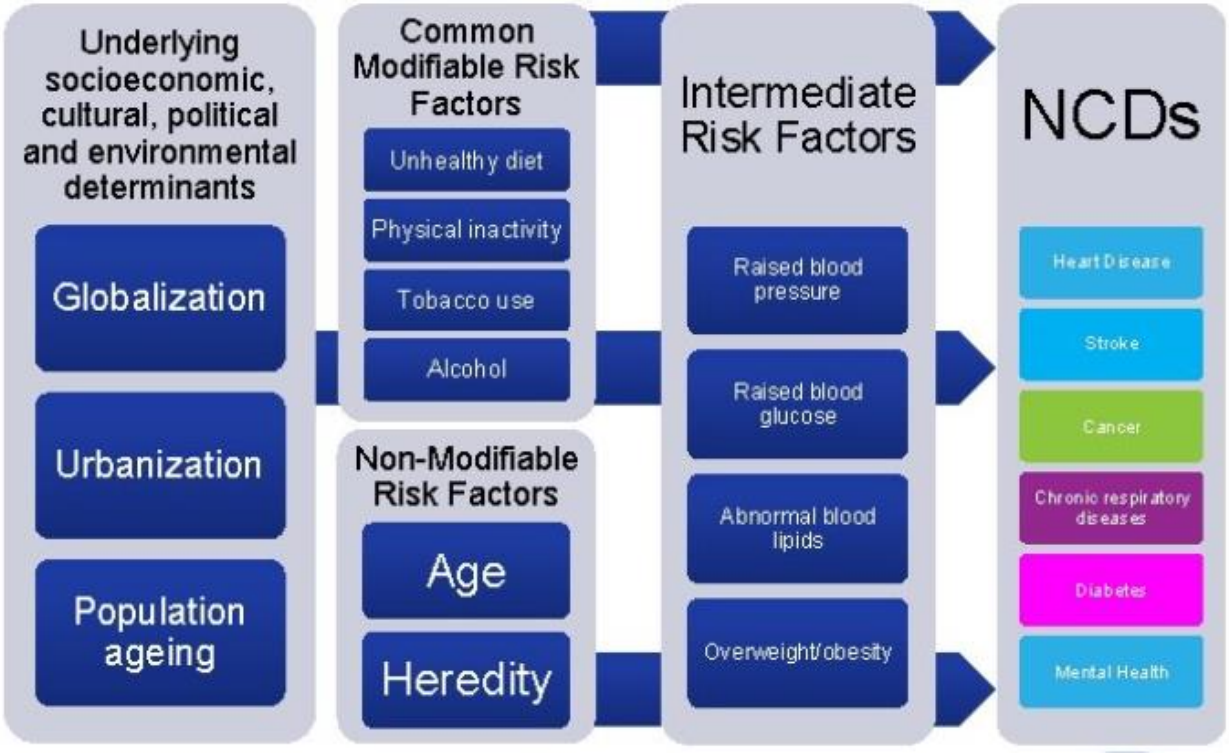


Figure 1. 2: The NCD causation pathway

[Adapted from WHO (2005) Preventing Chronic Disease: A Vital Investment]

1.5 Novel application of statistical methods

Despite the rapid increase in the use of latent class analysis (LCA), joint shared spatial statistics, and generalized structural equation modelling, to our knowledge, there is limited use of these methods to broaden our understanding of NCD clustering, and the spatial distribution and factors associated with MM. Applying these methods to MM research allows for (i) grouping participants according to the probability of disease co-occurrence, (ii) dividing further the unexplained variation (error term) of factors associated with MM into unmeasured shared components and an error term, and lastly, (iii) concurrently estimating the effects of several models that explore the association of different factors with MM, respectively. Further details of the methods are given in the next chapter and respective papers.

Psychological and medical literature uses LCA to model the relationships between variables in order to identify a group characterized by one or more latent variables [54-60]. For example, in recent years, LCA was applied to investigate alexithymia (the inability to recognise one's own emotions) in relation to other clinical psychiatric constructs. Among others, Vanheule and colleagues' study aimed to explore whether alexithymia is a stable trait that is independent from depression or a mental state secondary to depression [61]. In addition, Hartwig and colleagues examined the usefulness of a typological approach that considers the interaction between distinct alexithymic features within a population of high-alexithymic German adults, using a latent profile analysis [62-64].

Once the latent variables have been hypothesized, introducing additional external variables, such as gender, made it possible to investigate whether males and females differ in relation to the identified latent variables structure. LCA modelling is preferred over traditional clustering techniques such as distribution-based and hierarchical clustering. In LCA modelling, variation on observed indicators is modelled as a function of membership in unobserved classes called latent classes [65,66]. In addition, LCA allows for statistical testing of model fit and class membership in a probabilistic way, with membership probabilities computed from the estimated model parameters [64]. Furthermore, LCA has been demonstrated to be more objective and rigorous than K-means and hierarchical clustering for both exploratory work and theory testing [67]. This is because LCA is model based, i.e. there is a statistical model that is assumed to come from the population from which the data was gathered [64]. In the current study, seven chronic health conditions (angina, arthritis, asthma, chronic lung disease, depression, diabetes, and hypertension) were used as observed indicators. The optimal

number of latent classes was determined using the adjusted Bayesian Information Criterion (aBIC), which has been shown to provide robust indicators of class enumeration with categorical outcomes [68]. The adjusted BIC was used to compare several plausible class models where the lowest values indicate the best fitting model. After selecting the best model, each participant was assigned to one class according to his or her highest computed probability of membership. Given these premises, the present study mainly aims: (1) to investigate, by means of the LCA, if one or more latent variables exist which represent the latent structure of the relations among the manifest variables measured by the seven NCDs; and (2) to analyse if the external variables, such as demographic, socio-economic, behavioural variables, might be associated with the latent classes using multinomial logistic regression. The LCA is also preferred over other non-categorical methods, such as the structural equation modelling framework, after considering that important external variables may be categorical. Additionally, the observed variables can also be considered as categorical according to a discretization procedure applied to continuous data [69].

Joint disease mapping models are a direct extension of univariate spatial models that use both global and local spatial dependence structures to model risk of diseases. The extensions enable analysts to assess similarities as well as differences between risk factors for diseases which share common risk profiles [70-75]. One such joint disease mapping model is the shared component model which fits common and disease-specific unobserved and unmeasured spatial risks [71,72,76]. The shared component can be interpreted as a surrogate for unobserved covariates that display spatial structure and are common to both diseases. Similarly, each disease-specific component represents those spatially varying risk factors which are specific to the respective disease. The model explores the potential role of unmeasured characteristics in the co-occurrence of diseases (the shared component). Such multivariate models have been used for the following reasons, firstly, the correlation structures between relative risks of related diseases are implicitly quantified. Secondly, the models enable common and disease-specific observed covariate effects as well as spatial patterns to be estimated at the same time [76,77]. Similar work has used joint mapping models in cancer research, childhood illnesses, and childhood cancer research as well as diabetes research [72,73,78]. For instance, Manda and colleagues extended the spatio-temporal methodology by including shared spatial and temporal trends using a more extensive dataset among individuals diagnosed with acute lymphoblastic leukemia (ALL) and type 1 diabetes (T1D) in

Yorkshire (UK) aged 0–14 years from 1978–2003. A Bayesian model was fitted where similarities and differences in risk profiles of the two diseases were captured by the shared and disease-specific components using a shared component model, with space-time interactions. The shared component was identified as "environmental factors that exhibit similar geographical-temporal variation" [75].

Lastly, gSEM was utilized to explore the complex inter-relationships between socio-economic, socio-demographic, behavioural, anthropometric, and environmental factors, and how they are then associated with multimorbidity. gSEM methodology has been suggested as an alternative for modelling complex networks for a number of reasons. Firstly, gSEM models all equations simultaneously. This provides a flexible and general framework to test several potential relationships between a number of variables in the model [79]. Moreover, gSEM is preferred due to its capability to quantify each of the factors' contribution to the covariance structure. This caters for the limitation of linear models to control for confounding among continuous variables and interaction among categorical variables [80]. In addition, gSEM has the ability to simultaneously handle nested or crossed group-level effects in a particular data set i.e. latent and observed variables that may vary at different levels can be concurrently modelled [81].

1.6 Problem statement

There is a need for research in South Africa to assess the co-occurrence of NCDs particularly among middle aged and older adults using novel applications of advanced statistical techniques that could provide more insight. This is despite the ability of the tools to give more insight on the complex relationship that exist between MM, depression, and their associated factors. Moreover, there is need to identify MM co-occurrences in middle to older aged adult population for PHC care optimization. Lastly, evidence on determining the "hot-spots" of MM in SA for intervention purposes is sparce.

1.7 Justification

Associations have been found between socioeconomic disadvantage and poor disease control, an important consideration in South Africa where the majority of the population is dependent upon public sector health services [82]. The rising NCD burden necessitates an integrated approach to chronic disease care, including equipping primary health care providers to manage NCDs and the complexities of multimorbidity. The identification of hotspots of multimorbidity in South Africa using bivariate joint spatial shared modelling will potentially provide guidance to policy makers to implement more localised approaches in intervention targeting prevention, screening, and treating NCDs

rather than at national level. Further, the application of gSEM which allows for modelling several pathways concurrently in identifying factors associated with MM and depression will provide more precise model estimates and hence exhibiting the actual extent of the factors related to MM. Our study, to our knowledge, will be the first to utilize the gSEM model to examine the inter-relationships between different factors and how they are associated with multimorbidity in middle aged and older South African adults.

The South African health care system has adopted an integrated chronic disease management (ICDM) approach aimed at optimising the delivery of services for patients with chronic diseases [83]. However, implementation of this model has not been particularly successful and many programmes are still disease focused and still practice vertical implementation which does not consider comorbidities [84]. In addition, scale-up and sustainability of the approach have proven a problem [83]. Chronic diseases and risk factors are often undiagnosed and inadequately treated in South Africa, resulting in high levels of uncontrolled hypertension, diabetes, and chronic respiratory diseases [13,85,86]. Moreover, though it is established that NCDs are co-occurring (in our context depression, diabetes and hypertension) the South African health system continues largely treating them individually [84].

1.8 Aims and objectives

This thesis, therefore, aimed to examine the prevalence of NCD multimorbidity; to identify co-occurrence patterns and factors associated with MM in adult South Africans using latent class analysis; assess the patterns of the shared component among hypertension and diabetes using the bivariate joint spatial shared modelling; and the interrelationships of variables (demographic, environmental, economic, and behavioural) with multimorbidity using gSEM. This aim was addressed in three parts (objectives):

1. To determine the prevalence of multimorbidity in a cohort of South African adults over the age of 18 years, to identify patterns of chronic diseases co-occurrence, and to identify the demographic, anthropometric and behavioural factors associated with the different multimorbidity classes.
2. To examine the spatial distribution of hypertension and diabetes and the distribution of shared unmeasured characteristics on hypertension and diabetes in South African middle aged and older adults.

3. To determine the complex inter-relationships between socio-economic, demographic, behavioural, and environmental factors that are associated with multimorbidity and depression in South African middle aged and older adults.

Chapter 2 describes the study sample, variables, and statistical methods in detail.

CHAPTER 2: METHODS

To illustrate the novel application of the statistical techniques, data were obtained from the WHO's Study on global AGEing and adult health (SAGE) Wave-2 (2014/15) [9] in South Africa. SAGE is a panel study with nationally representative samples of persons aged 50 and above, with a comparative sample of individuals who are aged between 18 and 49 years, in China, Ghana, India, Mexico, the Russian Federation and South Africa [9]. Wave 2 is an implementation of the SAGE follow-up of Wave 1, 7 years later. For this thesis, we considered all participants successfully interviewed at the second wave only (N=3153) in South Africa. In the SAGE sample, Wave 1 and Wave 2 participants could not be linked due to a software programming error. Individuals who reported having any of the NCDs in Wave 1 were not asked about diagnosis information in Wave 2.

In the WHO SAGE study, respondents were asked to state whether they were diagnosed with one or more of the following chronic health conditions: angina pectoris, arthritis, asthma, chronic lung disease, depression, diabetes mellitus, and hypertension. In the analysis, these health conditions were combined to create a summative index of MM ranging from 0 to 7 for objective 1 and 2 and 0 to 6 for objective 3. A binary variable was then created to indicate: 0) no or one NCD and 1) presence of at least 2 NCDs.

Sociodemographic data that was collected included sex and age. Environmental factors were area of residence (rural/urban), and perception of safety in the local area (unsafe/moderate/safe). Socio-economic measures included a list of household assets and years of education. Behavioural variables included whether or not an individual added salt to food at table, history of tobacco use and history of drinking alcohol, sleep quality and exercise. Details on the measurement of these variables are given in detail in Chapter 3 through 5. Anthropometric measures included weight, height and waist circumference and were measured in accordance with WHO standardised techniques with all fieldwork teams trained by WHO staff. Body Mass Index (BMI; weight, kg / height, m^2), and waist to height ratio [waist (cm) / height (cm)] were calculated. Details about the WHO standardised interview and direct measurement techniques are described elsewhere [9].

2.1 Statistical methods

2.1.1 Latent class analysis

Latent class analysis was used in study 1 to classify South African adults according to MM risk, using self-reported diagnosed NCD health condition variables. Latent class analysis (LCA) involves the postulation of parameters in a statistical model, allowing the parameters to differ across unobserved subgroups. These subgroups form the categories of a categorical latent variable. This basic idea has several seemingly unrelated applications, the most important of which are clustering, scaling, density estimation, and random-effects modelling. Outside social sciences, LC models are often referred to as finite mixture models.

LC analysis was introduced in 1950 by Lazarsfeld [63], who used the technique as a tool for building typologies (or clustering) based on dichotomous observed variables . More than 20 years later, Goodman (1974) made the model applicable in practice by developing an algorithm for obtaining maximum likelihood estimates of the model parameters [65]. He also proposed extensions for polytomous manifest variables and multiple latent variables and did important work on the issue of model identification. During the same period, Haberman (1979) showed the connection between LC models and log-linear models for frequency tables with missing (unknown) cell counts [87]. Many important extensions of the classical LC model have been proposed since then, such as models containing (continuous) covariates, local dependencies, ordinal variables, several latent variables, and repeated measures. A general framework for categorical data analysis with discrete latent variables was proposed by Hagenaars and extended by Vermunt [88,89].

While in the social sciences LC and finite mixture models are conceived primarily as tools for categorical data analysis, they can be useful in several other areas as well. One of these is density estimation, in which one makes use of the fact that a complicated density can be approximated as a finite mixture of simpler densities. LC analysis can also be used as a probabilistic cluster analysis tool for continuous observed variables, an approach that offers many advantages over traditional cluster techniques such as K-means clustering. Another application area is dealing with unobserved heterogeneity, for example, in regression analysis with dependent observations.

Log-linear formulation of the LC model

Haberman (1979) showed that the LC model can also be specified as a log- linear model for a table with missing cell entries or, more precisely, as a model for the expanded table

including the latent variable *X* as an additional dimension [87]. The relevant log-linear model for *P* (*X* = *x*, **Y** = **y**) has the following form:

$$InP(X = x, Y = y) = \beta + \beta_x^X + \sum_{l=1}^{L} \beta_{y_l}^{Y_l} + \sum_{l=1}^{L} \beta_{x,y_l}^{X,Y_l} \text{.....2.1}$$

It contains a main effect, the one-variable terms for the latent variable and the indicators, and the two-variable terms involving *X* and each of the indicators. Note that the terms involving two or more manifest variables are omitted because of the local independence assumption.

The connection between the log-linear parameters and the conditional response probabilities is as follows:

$$P(Y_l = y_l | X = x) = \frac{\exp(\beta_{y_l}^{Y_l} + \beta_{x,y_l}^{X,Y_l})}{\sum_{r=1}^{D_l} \exp(\beta_r^{Y_l} + \beta_{x,r}^{X,Y_l})} \text{.....2.2}$$

This shows that the log-linear formulation amounts to specifying a logit model for each of the conditional response probabilities.

The type of LC formulation that is used becomes important if one wishes to impose restrictions. Although constraints on probabilities can sometimes be transformed into constraints on log-linear parameters and vice versa, there are many situations in which this is not possible.

Maximum likelihood estimation

Let *I* denote the total number of cells entries (or possible answer patterns) in the *L*-way frequency table, so that $I = \text{Q}L \quad D_A$, and let *i* denote a particular cell entry, n_i the observed frequency in cell *i*, and *P* (**Y** = $\mathbf{y}_i$) the probability of having the response pattern of cell *i*.

The parameters of LC models are typically estimated by means of maxi- mum likelihood (ML). The kernel of the log-likelihood function that is maximized equals:

$$InL = \sum_{i=1}^{I} n_i InP(Y = y_i) \text{.....2.3}$$

Notice that only non-zero observed cell entries contribute to the log-likelihood function, a feature that is exploited by several more efficient LC software packages that have been developed within the past few years.

The most popular methods for solving the ML estimation problem are the Expectation-Maximization (EM) and Newton-Raphson (NR) algorithms. EM is a very stable iterative method for ML estimation with incomplete data. NR is a faster procedure that, however, needs good starting values to converge. The latter method makes use of the matrix of second-order derivatives of the log-likelihood function, which is also needed

for obtaining standard errors of the model parameters.

Model selection issues

The goodness-of-fit of an estimated LC model is usually tested by the Pearson or the likelihood-ratio chi-squared statistic (see categorical data analysis). The latter is defined as:

$$L^2 = 2\sum_{i=1}^{I} n_i \, ln \frac{n_i}{N.P(Y=y_i)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.4$$

where *N* denotes the total sample size. As in log-linear analysis, the number of degrees of freedom (*df)* equals the number of cells in the frequency table minus one, $\prod_{l=1}^{L} D_l - 1$, minus the number of independent parameters. In an unrestricted LC model:

$$df = \prod_{l=1}^{L} D_l - C.\left[1 + \sum_{l=1}(D_l - 1)\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.5$$

Although it is no problem to estimate LC models with 10, 20, or 50 indicators, in such cases the frequency table may become very sparse and, as a result, asymptotic p values can nolonger be trusted. An elegant, but somewhat time-consuming, solution to this problem is to estimate the p values by parametric bootstrapping. Another option is to assess model fit in lower-order marginal tables, for example, in the two-way marginal tables.

It is not valid to compare models with *C* and *C*+1 classes by subtracting their L^2 and *df* values because this conditional test does not have an asymptotic chi-squared distribution. This means that alternative methods are required for comparing models with different numbers of classes. One popular method is the use of information criteria such as Akaike's Information Criteria (AIC) and means Bayesian Information Criteria (BIC).

Another more descriptive method is a measure for the proportion of total association accounted for by a *C*-class model, $[L^2(1)\ L^2(C)]/L^2(1)$, where the L^2 value of the one-class (independence) model, $L^2(1)$, is used as a measure of total association in the *L*-way frequency table.

Usually, we are not only interested in goodness-of-fit, but also in the performance of the modal classification rule [see equation (3)]. The estimated proportion of classification errors under modal classification equals:

$$E = \sum_{i=1}^{I} \frac{n_i}{N}\{1 - max[P(X = x|Y = y_i]\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.6$$

This number can be compared to the proportion of classification errors based on the

unconditional probabilities $P(X = x)$, yielding a reduction of errors measure λ:

$$\lambda = 1 - \frac{E}{max[P(X=x)]} \text{.....2.7}$$

The closer this nominal R^2-type measure is to one, the better the classification

Software

The first LC program, MLLSA, made available by Clifford Clogg in 1977, was limited to a relatively small number of nominal variables. Today's program can handle many more variables, as well as other scale types. For example, the LEM program provides a command language that can be used to specify a large variety of models for categorical data, including LC models [89]. Mplus is a command language based structural equation modelling package that implements some kinds of LC models, but not for nominal indicators. In contrast to these command language programs; Latent GOLD is a program with an SPSS-like user interface that is especially developed for LC analysis. It implements the most important types of LC models, deals with variables of different scale types, and extends the basic model to include covariates, local dependencies, several latent variables, and partially observed indicators.

2.1.2 Bivariate joint spatial modelling

Bivariate shared component modelling was used to assess the clustering and association between diabetes and hypertension in study 2, and to identify the shared component risk profile of hypertension and diabetes among older adults in South Africa.

Joint modelling

Univariate disease mapping has been common in many studies. Many diseases however share common risk factors. The joint modelling of two or more diseases across a geographical area to estimate relative risks is of both methodological and epidemiological importance. By pooling all the available data from different disease sources, there are gains in precision and efficiency of estimates especially in rare diseases [70]. Joint modelling of diseases, other than being useful in helping to identify disease specific risk factors, also provides estimates and inferences on the pairwise and cross-covariance between the risks of disease outcomes [70,78]. The joint modelling is usually initiated via the random effects by several possible approaches among them the multivariate normal distribution (MVN), the multivariate conditional autoregressive model (MCAR) and the multiple membership multiple classification (MMMC) approaches. The random effects can be decomposed into structured random effects which account for any unobserved covariates which vary spatially across the regions

and unstructured random effects which caters for the unobserved covariates that are inherent within the regions under study.

Suppose that y_{ij1} and y_{ij2} represents the disease 1 and 2 status respectively of individual j living in country i. We assume that the dependent variable y_{ijk} follows a Bernoulli distribution, i.e. $y_{ijk}|p_{ijk} \sim (p_{ijk})$. The unknown

$$h(p_{ij1}) = X^T \beta_1 + W^T{}_{ij1}\gamma_1 + u_{i1} + v_i \text{.....2.8, for disease 1}$$

and

$$h(p_{ij2}) = X^T \beta_2 + W^T{}_{ij2}\gamma_2 + u_{i2} + v_{i2} \text{.....2.9, for disease 2}$$

where the vector $X_{ijk} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp}), j$ contains p continuous predictors and $W_{ijk} = (w_{ij1}, w_{ij2}, \ldots, w_{ijr}), j$ contains r categorical predictors with the first component accounting for the intercept, u_i and v_i represents the unstructured and the structured random effects respectively. A bivariate model to measure risk for the two diseases can be imposed via u_{ik} and v_{ik} or both $u_i = (u_{i1}, u_{i2})^T$ and $v_i = (v_{i1}, v_{i2})^T$. The unstructured random effects $u_i = (u_{i1}, u_{i2})^T$ are assigned a bivariate normal distribution with covariance matrix,$\sum_u$ to allow for correlation between the disease risks, $u_i \sim MVN_2(0, \sum_u)$, where $\sum_{u11} = \sigma_{u1}^2$. Similarly, for the spatially structured terms $v_i = (v_{i1}, v_{i2})^T$, either the bivariate intrinsic conditional autoregressive model (ICAR) or the MMMC model is used. For the bivariate ICAR model, the structured terms are assigned a variate normal distribution $v_i \sim MVN_2(0, \sum_u)$, where $\overline{V}_i$ is the mean vector:

$$\overline{V}_i = \left(\sum_{i\epsilon\Theta_i} \frac{v_{i1}}{m_i}, \sum_{i\epsilon\Theta_i} \frac{v_{i2}}{m_i}\right)^T, \text{.....2.10}$$

where Θ_i is the set of neighbors m_i is the number of neighbors of area i and $\sum_v$ is the covariance matrix $v_i = (v_{i1}, v_{i2})^T$

The conditional variance for v_{i1} and v_{i2} respectively $\sum_{v11} = \sigma_{v1}^2/m_i$ and $\sum_{v11} = \sigma_{v1}^2/m_i$

MMMC models have been previously applied in spatial epidemiology [78]. For example, Manda and colleagues used two classifications: an area classification capturing the non-spatial variation (classification level 2) and a neighbor classification (classification level 3) to capture effects due to neighboring areas. They used the following notations where the superscript represents the classification levels: $b_i = \sum_{i \neq j} W_{ij} u_j^{(3)}$ where W_{ij} is the weighting factor that relates area i to each of the neighbor j in the neighborhood set Θ_i and $u_j^{(3)}$ is the effect of area j and on area i weighted by W_{ij} while the non-spatial random effects $u_{area(i)}^{(2)}$ are assigned

independent normal distribution $u_{area(i)}^{(2)} \sim N(0, \sigma_{u(2)}^2)$ and areas in classification set Θ_i have random effects $u_j^{(3)} \sim N(0, \sigma_{u(3)}^2)$. The standard choice of the weighting function is similar to that of the MCAR (CAR) model i.e $W_{ij} = 1\frac{1}{m_i}$ where m_i is the number of neighbors implying that the more the neighbors an area has the more precision is for that area. The difference between the MCAR and the MMMC is that for MCAR, the spatial correlation is achieved through a variance structure while for MMMC the spatial correlation is achieved through a multiple membership relationship and that the neighborhood random effects are not independent.

2.1.3 Structural equation modelling

For Study 3 generalized structural equation modelling (gSEM) was used to explore the association between socioeconomic, environmental, demographic, and behavioural factors, and MM in middle aged and older adults. Structural equation modelling (SEM) is a comprehensive statistical tool that is employed to represent, estimate and test a network of relationships between observed (measured) variables and latent variables [90]. The distinction between SEM and gSEM is that in SEM, the outcome variable is continuous, and the regression model is linear while gSEM accommodates a wide range of outcomes, i.e., the outcome variable may be continuous, binary, ordinal, or a count. Furthermore, non-linear link functions are allowed. In medical research, SEM has gained popularity as a powerful multi-variate analysis tool due to its capability to handle the investigation of both simple and complex causal models [91,80]. Therefore, gSEM allows for the inclusion of unobserved and observed effects for subjects, subjects within group, group within subgroups. Fitting gSEM in addition to fitting linear models, therefore, enabled us to get a better understanding of the complex inter-relationships between socio-economic, socio-demographic and environmental factors, both directly and indirectly. An *a priori* conceptual framework was used (See Figure 2) to show the hypothesized associations between the variables used in the model. From this framework, the application of gSEM was illustrated in Stata/IC version 16.1 in estimating the associations in the different socio-economic, socio-demographic and environmental pathways with multimorbidity.

Steps in gSEM

According to Lei Pui-Wa, SEM as a procedure includes the following steps [92,90]:

I. Specification of the model
II. Fitting the model
III. Evaluating the model
IV. Modifying the model
V. Interpreting and reporting the results.

I. *Specification of the model*

Based on literature and field of expertise model specification includes hypothesizing relationships among the variables that will be analyzed. The conceptualized model is often represented in graphical form (See Figure 4). The relationship among the variables analyzed may either be direct, indirect or non-directional. Covariance between variables is depicted by two-headed arrows while single-headed arrows depict a direct causal effect [92].

II. Model Estimation

Model estimation refers to the process of estimating identified parameters (regression estimates, variances and covariance among predictors) in the specified structural model. This estimation process can be done using one of three possible iterative procedures namely maximum likelihood, maximum likelihood with missing values and asymptotic distribution free. In this study, parameters were estimated by maximum likelihood.

III. Evaluating the model

The difference between the observed data and the hypothesized model is minimized by model estimation [93]. However, a dichotomous decision must be made on whether the proposed model is rejected or retained. This can be done objectively to assess whether the model under consideration fits the observed data through statistical model fit tests such as the Chi-squared test as well as calculating the Standardized Root Mean Square Residual (SRMR). This process is known as model evaluation.

IV. Modifying the model

Model modification includes the following: based on the outcome of the model fit tests, if the proposed SEM model does not fit the observed data, the researchers often re-specify the model. This improvement is substantively informed by literature to avoid theoretical modification [94].

V. Interpreting and reporting the results

When the SEM researcher is satisfied with the steps described above, the final step would be to interpret and report the estimates obtained from the model.

The SEM methodology has been widely used in behavioural sciences due to its generality and flexibility [95,90]. SEM, as an extension of general linear model (GLM) procedures, has been preferred over other GLM methods for a number of reasons that include its allowance for estimation of direct, indirect, total as well as path specific effects; the ability to quantify each factor's contribution to the covariance structure and understanding patterns of covariance among variables; exploration of potential mediators and latent confounders and its ability to model all the equations simultaneously [90]. Furthermore, SEM explicitly recognizes and measures measurement error and resolves collinearity by describing multiple measures using a latent variable. Lastly, SEM provides a convenient way to present complex relationships pictographically [90].

In Chapter 3, I present the application of the LCA methodology to classify South African adults aged 45 years and older according to multimorbidity risk, using self-reported diagnosed NCD health condition variables in a latent class analysis using data from the World Health Organization Study on global AGEing and adult health (WHO SAGE) South Africa Wave 2.

CHAPTER 3: IDENTIFYING CO-OCCURRENCE AND CLUSTERING

OF CHRONIC DISEASES USING LATENT CLASS ANALYSIS: CROSS-SECTIONAL FINDINGS FROM SAGE SOUTH AFRICA WAVE 2

Data from this chapter has been published in *Chidumwa Glory, Maposa Innocent, Corso Barbara, Minicuci Nadia, Kowal Paul, Micklesfield Lisa K, Ware Lisa J.* ***"Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2"*** *BMJ Open* ***2021****;11:e041604. doi: 10.1136/bmjopen-2020-041604.*

Authors' Contributions:

Conceptualization, G.C., I.M., L.J.W., P.K. and L.K.M.; methodology, G.C., N.M., B.C., and I.M.; software, STATA, OPENBUGS; formal analysis, G.C.; investigation, P.K.; resources, WHO.; data curation, WHO; writing—original draft preparation, G.C.; writing—review and editing, G.C.; visualization, G.C.; supervision, I.M., L.J.W., L.K.M.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

3.1 Introduction

Non-communicable diseases (NCDs) are the leading cause of mortality across the globe [10], and accounted for 73% of deaths in 2017 [96,97]. In developed countries, it is estimated that approximately 1 in every 4 adults experience multimorbidity, with half of older adults having 3 or more chronic conditions [98,19]. The prevalence of NCDs continues to increase in low- and middle-income countries (LMICs) including South Africa [10]. NCDs are responsible for 43% of deaths per year in South Africa, with most being premature deaths (deaths occurring before the age of 65 years) [39-41]. NCD-related deaths are predicted to increase substantially over the next few decades if measures are not taken to combat the upward trend in prevalence [10,13].

Within an individual, the co-existence of two or more chronic non-communicable, mental health or infectious diseases, of long duration (>three months), is referred to as multimorbidity [99,15]. Data from a 2015 South African primary health care survey across all age groups reported the prevalence of NCD multimorbidity, which included hypertension, diabetes, ischaemic heart disease, asthma, epilepsy, chronic obstructive pulmonary disease, osteoarthritis and respiratory infection, as 14.4% [10]. A study by Garin and colleagues aimed at identifying and describing multimorbidity patterns among adults older than 50 years in low-, middle-, and high-income countries, using data from the Collaborative Research on Ageing in Europe project and the World Health Organization's Study on Global Ageing and Adult Health Wave 1, found that South Africa had a higher prevalence (68%) of multimorbidity (having at least two NCDs) than Ghana (48%), India (58%) and China (45%) [97]. In addition, in a study by Afshar and colleagues to compare the prevalence of multimorbidity across 28 low and middle-income countries using the World Health Survey (2003), the prevalence of multimorbidity (2 chronic conditions or more) in South Africa was 21.6% among the 50 to 64 years age-group and 30.1% among those aged 65 years and older [31]. A study by Ayeni and colleagues aimed at profiling multimorbidity among 2,281 South African women of age 18 years and older, newly diagnosed with breast cancer, across two South African provinces [100]. They reported that 43.9% of the women met the definition of multimorbidity which included conditions such as hypertension, HIV infection and tuberculosis.

Evidence suggests that the factors associated with the rising prevalence of NCDs in South Africa include age, area of residence (urban or rural), tobacco use, insufficient physical activity and unhealthy diets [13]. A study by Weimann et al., investigated the association between socioeconomic disadvantage and multimorbidity in South Africa at two time points,

2008 and 2012, using the National Income Dynamics Study (NIDS). They showed that the risk for multimorbidity was doubled in urban residents relative to their rural counterparts, and respondents who were socioeconomically deprived had a two-fold increased risk of having multimorbidity compared to the less-deprived in both urban and rural areas [11].

Previous research on multimorbidity in South Africa has primarily used simple counts of chronic conditions. However, different combinations of diseases may affect a person's health and health care differently [101]. To account for these differences, disease combinations can be categorized according to their likelihood to cluster together, pathophysiological pathways or management plans, for example, hypertension and diabetes frequently occur together and may share common pathophysiological mechanisms [101,102]. The prevalence and patterns of multimorbidity have important implications for targeted healthcare services for prevention, diagnosis, treatment, and control.

The aim of this study was to classify South African adults aged 45 years and older according to multimorbidity risk, using self-reported diagnosed NCD health condition variables in a latent class analysis using data from the World Health Organization Study on global AGEing and adult health (WHO SAGE) South Africa Wave 2. Additionally, the analyses looked at sociodemographic, anthropometric and behavioural factors associated with identified patterns of multimorbidity. The findings of the current study will contribute to the evidence base on the epidemiology of multimorbidity in a large South African adult population.

3.2 Methods

Study Design and Participants

The current study used data from the WHO SAGE South Africa, which is part of an ongoing multi-country longitudinal study including China, Ghana, India, Mexico, and the Russian Federation, to examine the health and wellbeing of nationally representative adult populations aged 18+ years in over 42,000 participants, with an emphasis on populations aged 50+ years [9]. Further details are available on the WHO SAGE website (http://www.who.int/healthinfo/sage/en/). The current study is a cross-sectional analysis for the SAGE South Africa Wave 2 data collected in 2014/5 using participants (n=1,967), who had valid (not equal to zero) post-stratification weights, who were at least 45 years of age, with full data on the seven target NCDs.

Measures

Data on seven chronic conditions were collected via measurement and/or self-report. Noting hypertension is a common NCD risk factor, for the purposes of this analysis we categorized it as one of the seven conditions. As previously described, blood pressure was measured by trained nurses using wrist-worn blood pressure devices with positioning sensor (R6, Omron, Japan)[103]. Hypertension status was determined as a measured average systolic blood pressure (SBP) reading of ≥140 mmHg; and/or an average diastolic blood pressure (DBP) reading of ≥90 mmHg; and/or current use (within the last 2 weeks) of antihypertensive medication [104]. Participants reported whether they had ever received a medical diagnosis for angina, arthritis, asthma, chronic lung disease (emphysema or bronchitis, chronic obstructive pulmonary disease), depression and diabetes. These six self-reported NCDs were assessed through a question about ever being diagnosed with the disease by a physician/health professional. The specific question was, "Have you ever been told by a health professional/doctor that you have (disease name)?".

Demographic variables included age, sex, years of schooling completed, and area of residence (urban or rural). Behavioural variables included ever used alcohol, ever used tobacco (smoked and smokeless), adding salt at the table (yes/no), participation in self-reported vigorous intensity activity (yes/no – "*Does your work involve vigorous-intensity activity that causes large increases in breathing or heart rate, [like heavy lifting, digging or chopping wood] for at least 10 minutes continuously?", and "Do you do any vigorous intensity sports, fitness or recreational (leisure) activities that cause large increases in breathing or heart rate [like running or football], for at least 10 minutes continuously?"*,), and self-rated sleep quality (very good/good, moderate or poor/very poor) as reported previously [9]. Anthropometric measures included weight, height and waist circumference and were measured in accordance with WHO standardised techniques with all fieldwork teams trained by WHO staff. Details about the WHO standardised interview and direct measurement techniques are described elsewhere [9]. Body Mass Index (BMI; weight, kg / height, m^2), and waist to height ratio [waist (cm) / height (cm)] were calculated. Principal components analysis (PCA) was used to derive a socioeconomic status (SES) index for each household. PCA involved using household ownership of a set of 19 assets, household density and household service access (sanitation and electricity) into categorical or interval variables. The variables were then processed in order to obtain weights and principal components. The results obtained from the first principal component (explaining the most variability) were used to develop an index. The SES indices were then grouped into tertiles, reflecting different SES levels in the wealth continuum, as previously applied [105-107].

Statistical Analysis

Data were captured using an electronic data capture system (CAPI). STATA Statistical Software: Release 16.0 (Stata Corp LLC, 2017; College Station, USA) was used for statistical analyses. The latent class analysis (LCA) was performed in SAS PROC LCA add-on to determine patterns of co-existing chronic health conditions in the 1967 participants. LCA modelling is preferred over traditional clustering techniques as variation on observed indicators is modelled as a function of membership in unobserved classes called latent classes [65,66]. In addition, LCA allows for statistical testing of model fit and class membership in a probabilistic way, with membership probabilities computed from the estimated model parameters [64]. Furthermore, LCA has been demonstrated to be more objective and rigorous than K-means and hierarchical clustering for both exploratory work and theory testing [67]. This is because LCA is model based, i.e. there is a statistical model that is assumed to come from the population from which the data was gathered [64]. In the current study, seven chronic health conditions (angina, arthritis, asthma, chronic lung disease, depression, diabetes, and hypertension) were used as observed indicators. The optimal number of latent classes was determined using the adjusted Bayesian Information Criterion (aBIC), which has been shown to provide robust indicators of class enumeration with categorical outcomes [68]. The adjusted BIC was used to compare several plausible class models where the lowest values indicate the best fitting model. After selecting the best model, each participant was assigned to one class according to his or her highest computed probability of membership. Details for the latent class analysis fit statistics are given in supplementary table 1. The Pearson's Chi-square test was used to test statistical differences between latent classes and categorical variables. Due to non-normality of continuous data, as shown by the Shapiro-Wilk test, statistical differences between groups/classes on continuous outcomes were tested using the Kruskal-Wallis test. Multinomial logistic regression was used to determine which sociodemographic, anthropometric and behavioural factors were associated with observed latent class membership. In the current study, we used STATA terminology for multinomial logistic regression. Relative risk ratios' (RRR's), 95% confidence intervals (CI's), and p-values are reported for each explanatory variable.

Patient and public involvement

This study did not involve any patient and/or public.

3.3 Results

A total of 1,967 participants were included in this analysis. Figure 3.1 shows the study flow diagram.

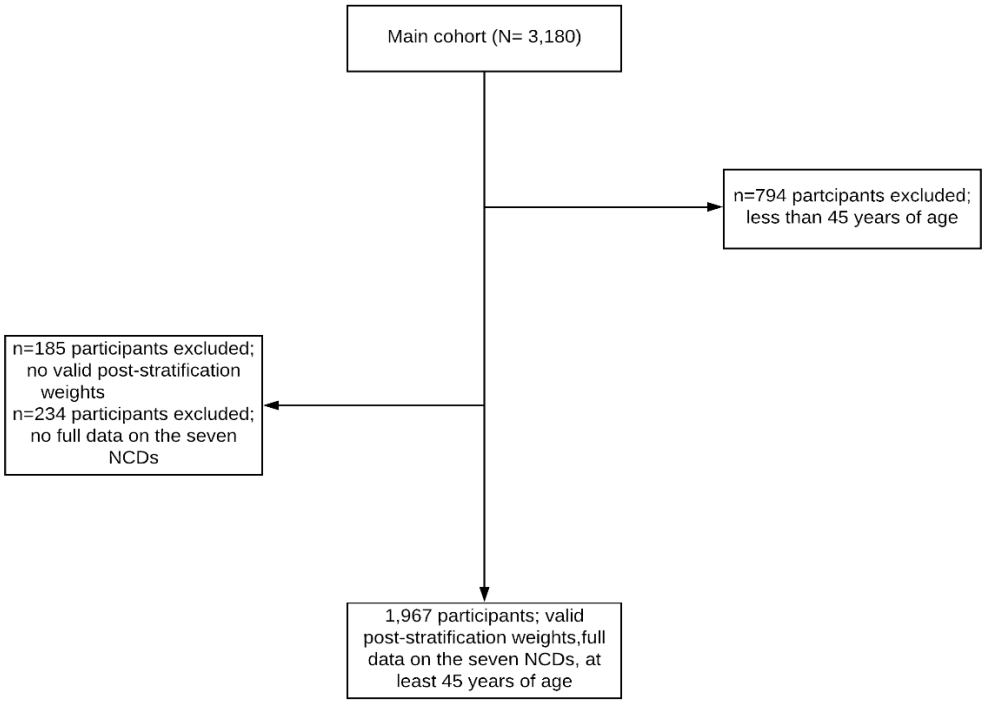


Figure 3. 1: Study flow diagram

The median age for the sample was 62 years [Inter-quartile range (IQR): 54 – 70]. Fifty-seven percent (n=1,113) of our sample were female. The majority of the sample self-identified as Black (n=1,540, 78%), 6% (n=120) as White, and 16% (n=308) as Coloured or Indian.

Prevalence of Chronic NCDs and Multimorbidity

Twenty-one percent of the sample (n=415) had two or more of the seven chronic diseases, i.e., multimorbidity (MM) while 39% (n=761) had none of the seven NCDs. The most common chronic disease was hypertension (52%) followed by arthritis (16%). Figure 3.2 shows the prevalence of chronic NCDs by sex.

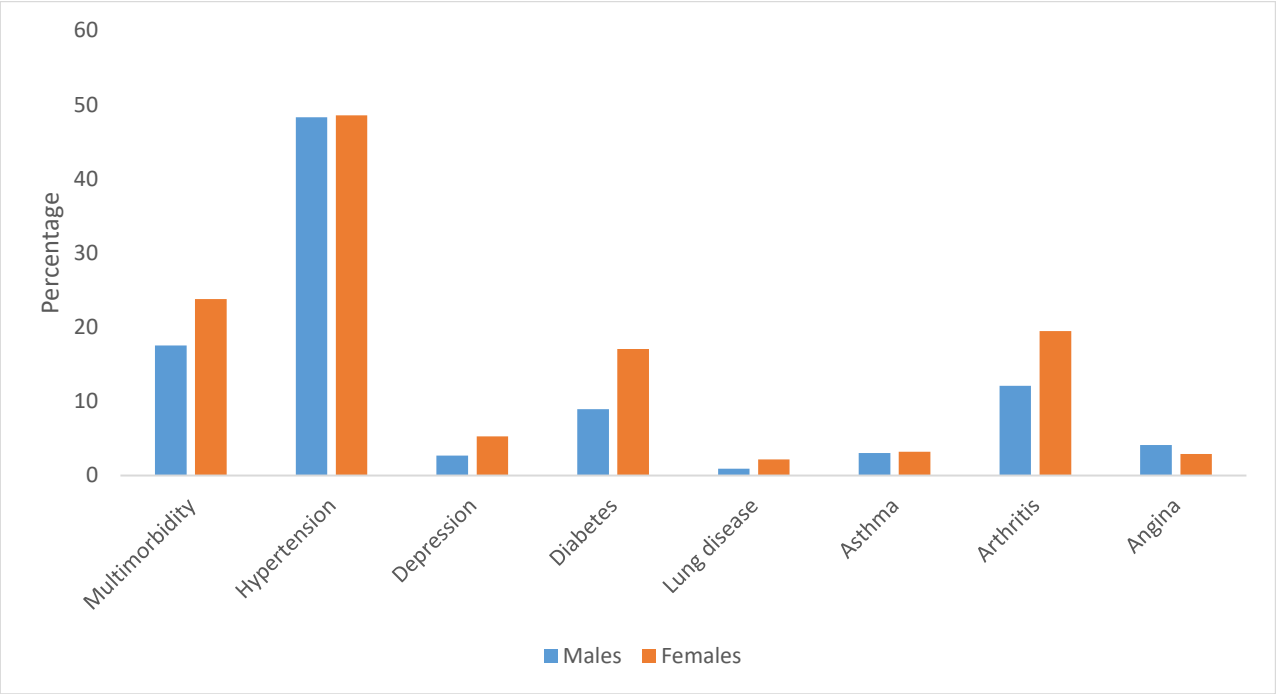


Figure 3. 2: Prevalence of chronic NCDs by sex

The prevalence of arthritis, depression, diabetes, lung disease and multimorbidity were higher in the women, and of angina were higher in the men.

Latent Classes for Chronic Disease Clusters

The optimal number of latent classes was determined using the adjusted BIC. There were negligible differences between the two class and three class models and considering plausible interpretability, the three-class model was chosen [68,108]. The three classes determined were: "*minimal MM risk*", which included the individuals with low probabilities for having each of the seven NCDs; "*concordant MM*" which included individuals with high probabilities of having hypertension and diabetes; and, "*discordant MM",* which included individuals with higher probabilities of having chronic conditions other than hypertension and diabetes. Concordant MM has been described by Piette and associates as chronic conditions that represent the similar pathophysiologic risk profile and are more likely to be the focus of the same disease management plan, and discordant MM as chronic conditions that are not directly related in pathogenesis or management [101]. The majority of the sample (n=1,625, 83%) were classified as being in the "minimal MM risk" class. This class had the lowest prevalence of all seven NCDs. The "*concordant MM*" class constituted 11% (n= 207) of the sample. The probability of being hypertensive in this class was 95%, and 74.1% for diabetes. Lastly, the "*discordant MM*" class comprised 6% (n= 135) of the sample, and showed prevalence of arthritis (62.0 %), angina (33.0%), asthma (11.7%), depression (15.3%), and

lung disease (34.1%). The prevalence of each of the seven diseases are presented by latent class as Supplementary Figure 3.1.

The demographic, anthropometric and behavioural characteristics of the three latent classes are presented in Table 3.1. The latent classes were significantly different with respect to all characteristics, except for self-reported vigorous intensity activity. Details of the pairwise comparisons between the groups are shown in Table 3.1.

Table 3. 1: Characteristics of participants by latent class category (n = 1967)

Variable	Minimal MM risk (N = 1591)	Concordant MM (N = 248)	Discordant MM (N = 128)	P-value
Age (years)	61 (54; 69) [a]	65 (58; 72) b	62 (55.5; 69) [a]	<0.001
BMI	28.5 (24.2; 34.4) [a]	29.5 (25.6; 35.6)	31.1 (25.2; 37.5) [b]	0.02
Waist circumference (cm)	94 (81; 105) [a]	99 (88; 109) [b]	100 (88; 112) [b]	<0.001
Hip circumference (cm)	100 (90; 112) [a]	106 (94; 116) [b]	106.5 (93; 118) [b]	<0.001
Waist to height ratio	0.579 (0.503; 0.662) [a]	0.642 (0.571; 0.728) [b]	0.634 (0.553; 0.710) [b]	<0.001
Years educated	8 (6; 11) [a]	8 (5; 10) [b]	8 (6; 10)	0.023
Sex				<0.001
Male	545 (34.3) [a]	51 (20.6) [b]	27 (21.1) [b]	
Female	1046 (65.7)	197 (79.4)	101 (78.9)	
Alcohol				0.033
Yes	289 (18.2) [a]	31 (12.7) [b]	29 (22.8) [a]	
No	1296 (81.8)	214 (87.3)	98 (77.2)	
Tobacco				<0.001
Yes	301 (19.0) [a]	35 (14.3) [a]	40 (31.7) [b]	
No	1284 (81.0)	210 (85.7)	86 (68.3)	
Add salt at table				0.013
Yes	1084 (68.4) [a]	155 (63.3)	73 (57.0) [b]	
No	501 (31.6)	90 (36.7)	55 (43.0)	
Self-reported vigorous intensity activity				0.325
Yes	181 (11.5)	26 (10.7)	20 (15.6)	
No	1396 (88.5)	218 (89.3)	108 (84.4)	
Residence				0.013
Urban	1124 (70.6) [a]	160 (64.5) [a]	101 (78.9) [b]	
Rural	467 (29.4)	88 (35.5)	27 (21.1)	
Household wealth tertile				0.001
1 (Lowest)	395 (80.58) [a]	39 (8.03) [a]	56 (11.39) [b]	
2	473 (80.96)	74 (12.61)	38 (6.42)	
3 (Highest)	455 (83.81)	58 (10.67)	30 (5.52)	
Sleep quality				<0.001
Good	1307 (83.2) [a]	176 (73.0) [b]	89 (71.2) [b]	
Bad	263 (16.8)	65 (27.0)	36 (28.8)	

[a-b] Medians in a row without a common superscript letter differ (P < 0.05), as analysed by the Dunn's multiple-comparison test for stochastic dominance using a Bonferroni correction for continuous data and pairwise Chi-square test for categorical data; P-values shown are for Kruskal Wallis test for continuous data and pairwise Chi-square test for categorical data. For categorical data frequencies are reported with percentages in parenthesis while medians are reported for continuous data with inter-quartile ranges in parenthesis.

Multinomial logistic regression results showing associations between the demographic, anthropometric and behavioural characteristics, and latent class membership, are presented in Table 3.2.

Table 3. 2: Results from multinomial logistic regression for factors associated with latent class membership

Reference (minimal MM risk)	Concordant MM		Discordant MM	
Characteristic	**Relative Risk Ratio (95% CI)**	**P-value**	**Relative Risk Ratio (95% CI)**	**P-value**
Age (years)	1.08 (1.04; 1.12)	<0.001	1.09 (1.04; 1.14)	0.001
Sex				
Male	Reference		Reference	
Female	4.38 (1.42; 13.6)	0.011	2.04 (0.58; 7.24)	0.267
Alcohol				
No	Reference		Reference	
Yes	1.13 (0.13; 9.76)	0.908	0.37 (0.08; 1.70)	0.201
Tobacco				
No	Reference		Reference	
Yes	2.92 (0.61; 13.9)	0.178	8.86 (2.03; 38.8)	0.004
Add salt at table				
No	Reference		Reference	
Yes	1.00 (0.43; 2.33)	0.992	0.53 (0.23; 1.22)	0.136
Physical activity				
No	Reference		Reference	
Yes	1.12 (0.48; 2.61)	0.784	0.77 (0.26; 2.30)	0.639
Residence				
Urban	Reference		Reference	
Rural	1.14 (0.41; 3.21)	0.799	1.31 (0.43; 4.00)	0.633
Household wealth tertile				
1 (Lowest)	Reference		Reference	
2	1.10 (0.45; 2.71)	0.833	0.61 (0.23; 1.58)	0.303
3 (Highest)	1.49 (0.38; 5.8)	0.564	0.43 (0.05; 3.75)	0.443
Sleep quality				
Good/Very good	Reference		Reference	
Moderate	1.58 (0.67; 3.72)	0.292	1.65 (0.57; 4.77)	0.350
Poor/Very poor	2.38 (0.66; 8.55)	0.183	0.99 (0.23; 4.34)	0.989
BMI	0.98 (0.92; 1.04)	0.54	1.01 (0.97; 1.06)	0.564
Years educated	1.00 (0.91; 1.11)	0.861	1.01 (0.83; 1.23)	0.939

In this multinomial logit model, we used the *minimal MM risk* group as the reference. Being female was associated with a 4.4-fold greater likelihood of being in the concordant group, and a one-year increase in age was associated with an 8% increased likelihood of being in the concordant group. Tobacco users were 8.9 times more likely to belong to the *discordant MM* class relative to the *minimal MM risk group.* Every year increase in age was significantly associated with a 9% increased likelihood of belonging to the *discordant MM* class. None of the other factors were significant in this logistic regression.

3.4 Discussion

In this study, we have shown that the prevalence of multimorbidity (co-existence of two or more NCDs) was 21%. The latent class analysis grouped our sample of men and women over the age of 45 years into three groups namely: *minimal MM risk* (83%), *concordant MM* (11%) and *discordant MM* (6%). When compared to the *minimal MM risk* group, being female and older were associated with belonging to the *concordant MM* group, while tobacco use and an increase in age were associated with belonging to the *discordant MM* group.

Several recent studies have explored multimorbidity in South Africa [109,11,100,110], however this study has used data from the SAGE which represents the 50+ years South African population, to identify patterns of chronic disease co-existence. In addition, to our knowledge this is the first study in South Africa to use latent class analysis to identify patterns of chronic disease co-existence as LCA has the ability to identify unique combinations of diseases using probabilities [64].

Our study identified three latent classes of multimorbidity based on the presence or absence of seven chronic conditions. Previous studies that have used the LCA method to describe patterns of chronic disease co-existence in older populations have yielded mixed results as regards the number of clusters identified. In a cross-sectional sample of 4,574 Australian senior citizens (aged 50 years and over) using eleven chronic conditions, reported a MM prevalence of 52% and identified four classes [111]. Their sample presented (i) a relatively healthier group (ii) a sick group with dominant presence of arthritis, asthma and depression, (iii) a sick group with dominant presence of hypertension and diabetes and (iv) the sickest group with dominant presence of cancer, heart and stroke [111]. Similarly, a retrospective cohort study on 13 self-reported conditions from 14,502 Americans (65 years old and older) identified six classes using the LCA approach, and reported a MM of 67.3% [112]. The classes included: minimal disease class (prevalence of all conditions is below cohort average), nonvascular class (excess prevalence in cancer, osteoporosis, arthritis, arrhythmia, chronic obstructive pulmonary disease, psychiatric disorders), vascular class (excess prevalence in hypertension, diabetes mellitus, stroke), cardio-stroke-cancer class (excess prevalence in congestive heart failure, coronary heart disease, arrhythmia, stroke, and to a lesser extent hypertension, diabetes mellitus, cancer), major neurological disease class (excess prevalence in Alzheimer's disease, Parkinson's disease, psychiatric disorders), and very sick class (above average prevalence of all 13 conditions) [112]. Comparison with these studies is difficult

since the results might be influenced by the number and type of diseases included in the analysis, the characteristics of the sample, or how data on diseases were collected.

In our study we identified a class representing "minimal MM risk" (participants with low observed probabilities for the NCDs reported), which has previously been reported in other studies which also conducted LCA [108,111,112]. However, the prevalence of 83% classified as "*minimal MM risk*" in our study is larger than that described in these studies. This difference could be explained by the age of participants included in a study. For example, a study conducted by Olaya and colleagues which found that 63.8 % of their sample were classified in the minimal disease category had a mean age of 66 years while the average age in our study is 62 years [108]. This is further supported by our finding that the probability of MM increases with age.

In addition, we identified two more classes namely *concordant MM* and *discordant MM*. This is similar to the study conducted by Chang and colleagues in rural South Africa where they defined concordant conditions as cardio-metabolic conditions (hypertension, diabetes and angina), and discordant conditions as mental health illness, alcohol dependence and HIV infection [109]. Differences in the conditions in the discordant class could be attributed to the fact that the studies did not consider the same conditions except depression.

To provide better care for individuals with comorbid conditions, South Africa implemented the integrated chronic disease management (ICDM) plan in 2014 for primary health care [113]. However, evidence suggests that implementation has faced challenges with many programmes remaining disease focused and with vertical implementation that fails to consider comorbid conditions [83,84]. Our findings have the potential to guide policy in refining implementation of strategies to address ICDM, for example, targeting to address hypertension and diabetes together.

In addition, in keeping with previous literature, we found tobacco users to have a higher probability of discordant MM which included lung disease, asthma, arthritis and angina, compared to non-tobacco users [114-116]. For example, in a study by Fonda and colleagues aimed at examining the clustering of post-traumatic stress disorder, depressive disorders, and clinically significant pain among 433 deployed veterans in Boston (USA) aged 18 to 65 years, tobacco smokers had 3.5 increased likelihood for multimorbidity [117].

The findings from this study should be viewed considering some limitations. Firstly, since the current study design is cross-sectional in nature, we could not determine the temporal

sequence or causality. Second, data on most of the chronic diseases, and many behavioural variables (including tobacco use), was based on self-report, and can thus be affected by possible recall bias and social desirability bias. In addition, the definitions of alcohol use and tobacco use in our study were broad and do not capture the quantities and frequency of consumption, potentially explaining the lack of association found. Furthermore, the LCA combined participants without NCDs with those with mostly one NCD in the *minimal MM risk* group, thereby limiting the use of participants with no MM as the reference group. In addition, the LCA procedure was explorative in nature. Explorative LCA makes no priori assumptions about the number of latent classes and estimated starting with a two-class model and increasing the number of latent classes in a stepwise fashion. As such, when different criterions to determine the classes are used, researchers may argue in favour of different numbers of classes. Finally, the number of diseases included in this analysis was limited to those included in the SAGE study. This may miss other conditions present in this population, such as dementia or cancers, and therefore have resulted in an underestimation of multimorbidity prevalence. However, our prevalence data for MM is similar overall to previous SAGE recent data, and a number of studies have also analysed multimorbidity using a smaller number of diseases, usually less than 10, due to data collection limitations in LMICs such as lack of electronic health/medical records [110].

In conclusion, this study identified three latent classes namely: *minimal MM risk*, *concordant MM* and *discordant MM*. Review of the South Africa literature highlights that the primary health (PHC) system under the ICDM model remains single disease focused on the treatment of patients. In improving PHC in South Africa, efforts should be made to manage multiple conditions concurrently at PHC centers, in particular diabetes and hypertension. In addition, in our sample, risk factors for multimorbidity latent classes include age, sex and tobacco use. Future efforts should focus on the inclusion of all frequently occurring common conditions, including infectious diseases to evaluate clustering patterns and inform policymakers to prioritise the older population, women and tobacco users in prevention programmes.

Given the co-occurrence of hypertension and diabetes, the Chapter 4 describes in detail the spatial distribution of hypertension and diabetes. Moreover, these conditions are known to share common socio-demographic, anthropometric and lifestyle risk factors. Therefore, the shared component can be interpreted as a proxy for unobserved covariates that display spatial structure and are common to both hypertension and diabetes.

CHAPTER 4: BIVARIATE JOINT SPATIAL MODELLING TO IDENTIFY SHARED RISK PATTERNS OF HYPERTENSION AND DIABETES IN SOUTH AFRICA: EVIDENCE FROM WHO SAGE SOUTH AFRICA WAVE 2

Data from this chapter has been published in *Chidumwa Glory, Maposa Innocent, Kowal Paul, Micklesfield Lisa K., and Ware Lisa J.* ***2021****. "****Bivariate Joint Spatial Modelling to Identify Shared Risk Patterns of Hypertension and Diabetes in South Africa: Evidence from WHO SAGE South Africa Wave 2****" International Journal of Environmental Research and Public Health 18, no. 1: 359. https://doi.org/10.3390/ijerph18010359*

Author Contributions

Conceptualization, G.C., I.M., L.J.W., P.K. and L.K.M.; methodology, G.C. and I.M.; software, STATA, OPENBUGS; formal analysis, G.C.; investigation, P.K.; resources, WHO.; data curation, WHO; writing—original draft preparation, G.C.; writing—review and editing, G.C.; visualization, G.C.; supervision, I.M., L.J.W., L.K.M.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

4.1 Introduction

The global burden due to non-communicable diseases (NCDs) is high and rising; and is expected to increase in the next decades if public health interventions are not implemented to reduce the trend [10,13,5]. In sub-Saharan Africa, the total number of disability-adjusted life years (DALYs) due to NCDs for all ages increased by 67% between 1990 and 2017 [118]. Among African countries, this epidemiological transition has been reported to be attributable to the changing lifestyle patterns such as declining levels of physical activity [119]. In South Africa, NCDs have become the leading cause of mortality accounting for 40% of total deaths, with one-third of the deaths occurring before the age of 60 [120]. Recent studies have further highlighted the co-occurrence of NCDs. Oni and colleagues have highlighted the co-existence of multiple infectious diseases and NCDs in Cape Town adults from an informal settlement [121]. Their findings showed a 23% prevalence of multimorbidity (defined as having more than one chronic condition) among chronic disease patients, and patterns of multimorbidity with hypertension and diabetes often co-occurring [121]. In addition, in South Africa, non-diabetics with elevated blood pressure are 2.5 times more likely to develop diabetes within 5 years than individuals with normal blood pressure levels [122]. Individuals living with diabetes are twice as likely to have hypertension [104].

Data across multiple conditions may be pooled in a unified way using joint mapping models to better understand the overlapping epidemiology of the conditions. The present study used a bivariate spatial disease model to analyze hypertension and diabetes simultaneously [70,76,77]. Such multivariate models have been used for the following reasons, Firstly, the correlation structures between relative risks of related diseases are implicitly quantified. Secondly, the models enable common and disease-specific observed covariate effects as well as spatial patterns at the same time [76,77]. Similar work has used joint mapping models in cancer research, childhood illnesses, and childhood cancer research as well as diabetes research [72,73,75].

The current study aims to assess the shared component risk profile for hypertension and diabetes using data from the World Health Organization (WHO) Study on Global Ageing and Adult Health (SAGE) South Africa Wave 2. These conditions are known to share common socio-demographic, anthropometric and lifestyle risk factors [104,123]. Therefore, the shared component can be interpreted as a proxy for unobserved covariates that display spatial structure and are common to both diseases. Similarly, each disease-specific component represents spatially varying risk factors which are specific to the respective disease.

4.2 Materials and Methods

In this study, we used data from WHO SAGE South Africa Wave 2. SAGE is an ongoing multi-country longitudinal study that has also been implemented in China, Ghana, India, Mexico, and the Russian Federation. SAGE aims to examine the health and wellbeing of nationally representative adult populations aged 18+ years with an emphasis on populations aged 50+ years [9]. Further details are available at (http://www.who.int/healthinfo/sage/en/). In South Africa, 600 enumeration areas, with 30 households in each, were sampled from 18,000 targeted households. In the sample of households with people aged 50 years or older, all adults aged 50 years or older were eligible for interview. SAGE South Africa wave 1 total sample size for South Africa was 4223. Wave 2 is an implementation of the SAGE follow-up of Wave 1, 7 years later. In this sample, approximately 30% of participants interviewed in Wave 1 were interviewed again at Wave 2 and were linked using their unique identifiers. The remainder were new participants. The current analysis consists of 2761 participants who had valid (not equal to zero) post-stratification weights, with full data on hypertension and diabetes. The provincial samples of the participants who were 50 years and older in our study were representative of the middle-aged and older population in each province as shown in Supplementary Table 4.S1.

The unit of analysis for the current spatial model is district. The SAGE study used province and residence as the main stratification levels. South Africa consists of three structures of government—national, provincial, and local governments—and is divided into nine provinces, each with a provincial legislature (see Figure 4.1). The nine provinces are further divided into 52 districts. Provincial governments are bound by laws and policies passed at national level. However, provincial governments can adapt or develop their own laws and policies within the national framework to suit their specific needs [124]. The National Health Act requires provincial Departments of Health to develop their own strategic plans, which must conform with national health policy [125]. The provincial governments, therefore, implement their own priorities and allocate resources responsive to the needs of their populations.



Figure 4. 1: The South African map

[Source: https://en.wikipedia.org/wiki/Provinces_of_South_Africa].

Gauteng province is the most densely populated province with nearly eight hundred people per square kilometre, followed by Kwa-Zulu Natal province (178 people per square kilometre). The least densely populated province is the Northern Cape, with on three people per square kilometre. Gauteng contributes to approximately a third (34%) of South Africa's growth domestic product. The Eastern Cape and Limpopo provinces have the highest percentages of households in poverty, 12.7% and 11.5%, respectively, while Western Cape has only 2.7% households in poverty.

Outcome variables: Hypertension status was determined as a measured average (for three sequential readings) systolic blood pressure (SBP) readings of ≥140 mmHg and/or an average diastolic blood pressure (DBP) reading of ≥90 mmHg and/or self-reported hypertension with current use (within the last 2 weeks) of antihypertensive medication[104]. For the present

descriptive analysis, participants with systolic and diastolic blood pressure (BP) values <120/80 mmHg were classified as normotensive, while those with systolic BP from 120 to 139 mm Hg and diastolic BP from 80 to 89 mm Hg as prehypertensive. Self-reported diabetes status was assessed with the question "Have you ever been told by a health professional/doctor that you have diabetes?".

Exposure and predictor variables: Demographic variables included age, sex, years of schooling completed, and area of residence (urban or rural). Behavioural and social variables included ever used alcohol, ever used tobacco (smoked and smokeless), adding salt at the table (yes/no), self-reported vigorous intensity physical activity (yes/no—both leisure and work), and current employment status [9]. Anthropometric measures included waist circumference, weight and height; and were measured in accordance with WHO standardised techniques with all fieldwork teams trained by WHO staff. Waist to height ratio [waist (cm)/height (cm)] and body mass index were calculated [weight (kg)/height $(m)^2$). For descriptive purposes, we classified participants into the categories 'underweight' (body mass index (BMI) < 18.5 kg/m^2), 'normal weight' (BMI > 18.5 and <25.0 kg/m^2), 'overweight' (BMI > 25 kg/m^2) or 'obese' (BMI > 30 kg/m^2). Details about the WHO standardised interview and direct measurement techniques are described elsewhere [9]. Principal components analysis (PCA) was used to derive a socioeconomic status (SES) index for each household. PCA involved using household ownership of a set of 19 assets, household density and household service access (sanitation and electricity) into categorical or interval variables. The variables were then processed in order to obtain weights and principal components. The results obtained from the first principal component (explaining the most variability) were used to develop an index. The SES indices were then grouped into household wealth tertiles, reflecting different SES levels in the wealth continuum, as previously applied [105,106]. Financial support from government (yes/no) was also determined.

Statistical Methods

For the analyses, we used the following versions of software and packages: Stata Release 16.1 (Stata Corp LLC, 2017; College Station, TX, USA) and OpenBUGS version 3.2.3 rev 1012 [126]. Individual level data summary statistics, and provincial level prevalence of hypertension and diabetes were determined. In order to counter any possible negative confounding, individual factors to include in the final spatial shared models were selected a priori and using stepwise backward elimination multiple logistic regression for hypertension and diabetes separately ($p = 0.1$). The current analysis incorporates an ecological investigation to assess disease risk in relation to risk at individual level exposure. Joint disease mapping models are a

direct extension of univariate spatial models that use both global and local spatial dependence structures to model risk of diseases. The extensions enable analysts to make an assessment on similarities as well as differences between risk factors for diseases which share common risk profiles [70-75]. One such joint disease mapping model is the shared-component model which fits common and disease-specific unobserved and unmeasured spatial risks [71,72,76]. A detailed description of the shared-component model and its implementation within the Bayesian estimation procedure for hypertension and diabetes is given below:

Each j district (j = 1, …, 52) has N_j adults out of the total sampled, i.e., the sample of adults in the study from the total district population. We assume that X_{ij} is the vector of observed risk covariates associated with a subject i (i=1,..,2761) in district j. For the unaccounted variation in the risks of hypertension and diabetes, unobserved district-spatial variation U_{jk} is introduced for district j and NCD k (k = 1 and 2). We worked within the framework of conditional models where conditional on spatial random effects $U = (U_{j1}\ and\ U_{j2})$ and the disease-specific fixed effects parameters β_k, the binary responses Y_{ijk} were independent Bernoulli random variables with parameters π_{ijk}, being the probability of subject ij having disease k. In order to model the probabilities of the observed and unobserved spatial variation, we used a logit link function on the probabilities (i.e., joint spatial model without the shared component):

$$\log(\pi_{ijk}/(1-\pi_{ijk})) = \alpha_k + \beta_k' X_{ij} + U_{jk} + \varepsilon_j \qquad (4.1)$$

,where $\alpha_k{}'s$ are the disease-specific log-odds constant terms.

In order to model hypertension and diabetes together in a multivariate space, we used a shared-component model with one shared component, relevant to hypertension and diabetes. The shared spatial component could be interpreted as a proxy for variations in latent and unmeasured health behavior characteristics or the social determinants of health in our population. The known behavioral risk factors in our study include alcohol use, tobacco use, and adding salt at the table. We compared shared spatial component model with a spatial joint model to assess whether this model was better in capturing the underlying covariance structure of the data using the deviance information criteria (DIC). Within the symmetric formulations of the shared-component model, we also included disease-specific spatial components for hypertension and diabetes. Thus, the model decomposed each of the two spatial random effects $U_{j1}\ and\ U_{j2}$ into a common spatial and disease-specific component. The resulting model enables us to determine the extent of the variation exhibited through common as well as specific geographical patterns in the risks. We also allow for disease-specific unstructured

heterogeneous effects ϵ_{jk} to account for possible extra-binomial variation that was not explained by the included fixed effect, and common and specific structured spatial terms. Thus, hypertension and diabetes were modelled as follows on log-odds scale (i.e., joint spatial model with the shared component), an extension of Equation (1):

$$\log(\pi_{ij1}/(1-\pi_{ij1})) = \alpha_1 + \beta_1' X_{ij} + \gamma_1 Uj + U_{j1} + \varepsilon_{j1} \quad (4.2)$$

$$\log(\pi_{ij2}/(1-\pi_{ij2})) = \alpha_2 + \beta_2' X_{ij} + \gamma_2 Uj + U_{j2} + \varepsilon_{j2} \quad (4.3)$$

where $U_{j1}\ and\ U_{j2}$ are the log-odds structured random effects for each of hypertension and diabetes, in district j. γ_i's represent the risk gradient. The parameters $\alpha_k's$ and β_k' 's are the disease-specific baseline risk and fixed effect risks associated with the risk vector X_{ij}; and U_j is the shared component common to hypertension and diabetes.

4.3 Results

A total of 2761 adults provided the full set of health variables for our analysis, of which 641 (23.2%) had hypertension and 338 (12.3%) had diabetes. The median age was 56 years (inter-quartile range: 40–66 years). Mpumalanga province and Western Cape had the highest prevalence of hypertension (33.8% and 31.2%, respectively), while approximately one in five people are diabetic in Kwa-Zulu Natal and Western cape. Approximately 10% (n = 240) of our sample had comorbidity, defined as having both hypertension and diabetes. Comorbidity was significantly associated with demographic (age, sex, and current employment), socioeconomic (years of schooling, high school completion, household wealth tertile, receiving government support), anthropometric (waist-to-height ratio, BMI), and behavioral characteristics (adding salt at table). A detailed summary of the individual-level data and bivariate analyses are shown in Table 4.1 and 4.2.

Table 4. 1: Prevalence of hypertension, diabetes, and comorbidity (having both hypertension and diabetes or not), by province (N = 2761).

PROVINCE	Total	Hypertension	Diabetes	Comorbidity
	N	N (%)	N (%)	N (%)
Eastern Cape	522	82 (15.7)	31 (5.9)	18 (3.5)
Free State	216	51 (23.6)	22 (10.2)	13 (6.0)
Gauteng	528	117 (22.2)	80 (15.2)	55 (10.4)
Kwa-Zulu Natal	450	127 (28.2)	86 (19.1)	65 (14.4)
Mpumalanga	142	48 (33.8)	14 (9.9)	13 (9.2)
North West	318	68 (21.4)	26 (8.2)	14 (4.4)
Northern Cape	93	19 (20.4)	10 (10.8)	7 (7.5)
Northern Province	175	30 (17.1)	8 (4.6)	6 (3.4)
Western Cape	317	99 (31.2)	61 (19.3)	49 (15.5)

All values are frequencies and percentages in parenthesis.

Table 4. 2: Summary statistics for demographic, socioeconomic status, anthropometry, behavioral and blood pressure characteristics, by comorbidity, N = 2761.

	Total (N = 2761)	**No Co-Morbidity (N = 2521)**	**Co-Morbidity (N = 240)**	***p*-Value**
Demographic characteristics				
Age (years)				**<0.001**
Median (Q1, Q3)	56.0 (40.0, 66.0)	55.0 (39.0, 65.0)	64.0 (56.0, 71.0)	
Sex				**<0.001**
Male	915 (33.1)	866 (34.4)	49 (20.4)	
Female	1846 (66.9)	1655 (65.6)	191 (79.6)	
Currently working				**<0.001**
Yes	532 (34.5)	509 (36.4)	23 (16.4)	
No	1008 (65.5)	891 (63.6)	117 (83.6)	
Residence				0.611
Urban	1881 (68.1)	1721 (68.3)	160 (66.7)	
Rural	880 (31.9)	800 (31.7)	80 (33.3)	
Socioeconomic characteristics				
Years of schooling				**<0.001**
Median (Q1, Q3)	10.0 (7.0, 12.0)	10.0 (7.0, 12.0)	8.0 (6.0, 10.0)	
Completed high school?				**0.003**
Yes	652 (29.5)	615 (30.3)	37 (19.9)	
No	1561 (70.5)	1412 (69.7)	149 (80.1)	
Household wealth tertile				**0.008**
1 (lowest)	751 (33.4)	665 (32.6)	86 (41.3)	
2	749 (33.3)	678 (33.2)	71 (34.1)	
3 (highest)	748 (33.3)	697 (34.2)	51 (24.5)	
Support from government?				**<0.001**
Yes	936 (34.7)	827 (33.5)	109 (47.6)	
No	1764 (65.3)	1644 (66.5)	120 (52.4)	
Anthropometric characteristics				
Waist-to-height ratio				**<0.001**
Median (Q1, Q3)	0.6 (0.5, 0.7)	0.6 (0.5, 0.7)	0.6 (0.6, 0.7)	
Body mass index (BMI) category, kg/m^2				**0.012**
Underweight (<18.5)	60 (3.2)	57 (3.3)	3 (1.9)	
Normal weight (18.5–24.9)	494 (26.1)	467 (27.0)	27 (16.7)	
Overweight (25–29.9)	538 (28.5)	490 (28.3)	48 (29.6)	
Obese (≥30)	799 (42.3)	715 (41.4)	84 (51.9)	
Normotensive (<120/80 mmHg)				0.065
Yes	1342 (48.6)	1239 (49.1)	103 (42.9)	
No	1419 (51.4)	1282 (50.9)	137 (57.1)	
Pre-hypertensive (120/80-139/89mmHg)				0.466
Yes	1385 (50.2)	1270 (50.4)	115 (47.9)	
No	1376 (49.8)	1251 (49.6)	125 (52.1)	
Hypertensive				**<0.001**
Yes	1168 (42.3)	1041 (41.3)	127 (52.9)	
No	1593 (57.7)	1480 (58.7)	113 (47.1)	
Behavioral characteristics				
Add salt at table				**<0.001**
Yes	1910 (69.4)	1770 (70.4)	140 (58.8)	
No	842 (30.6)	744 (29.6)	98 (41.2)	
Self-reported vigorous intensity physical activity				0.468
Yes	366 (13.3)	338 (13.5)	28 (11.8)	
No	2376 (86.7)	2167 (86.5)	209 (88.2)	
Ever used alcohol?				0.221
Yes	523 (19.0)	485 (19.3)	38 (16.0)	
No	2227 (81.0)	2028 (80.7)	199 (84.0)	
Ever used tobacco?				0.781
Yes	482 (17.5)	442 (17.6)	40 (16.9)	
No	2267 (82.5)	2070 (82.4)	197 (83.1)	

Frequencies and percentages in parenthesis are shown for categorical data. *p*-values shown are for Mann–Whitney U test for continuous data (normality test checked using the Shapiro–Wilk test); and chi-square test for categorical data, and Fisher's exact test for BMI.

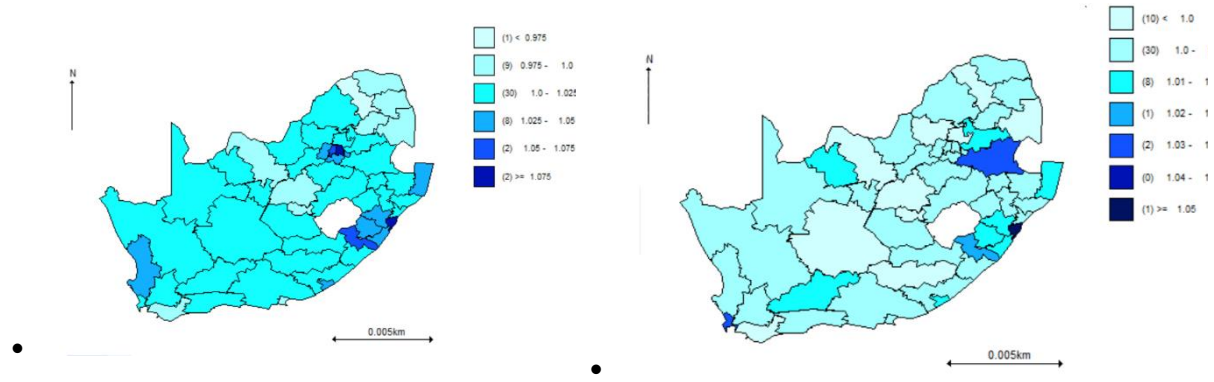
Results from the stepwise backward elimination multivariable analyses for hypertension and diabetes are shown in Supplementary Table 4.S2. Increasing age was positively associated with increased risks of both hypertension and diabetes. Being unemployed and having a few years of education were associated with higher hypertension and diabetes risks, although employment was not statistically significant. In addition, higher waist-to-height ratio was positively associated with increased risks of both hypertension and diabetes. Participants who had self-reported diabetes and depression had 84% and 66% higher risk of hypertension relative to their controls, respectively. Table 3 shows the multivariate logistic model, adjusting for demographic, socioeconomic status, anthropometry, behavioral characteristics, and spatial effects. Age was positively associated with increased odds of both hypertension and diabetes. Being in the lower household wealth tertile and having fewer years of schooling were associated with higher hypertension and diabetes risks. In addition, receiving support from government, adding salt to food at the table and having a higher BMI were associated with greater odds of hypertension. Moreover, tobacco use was associated with increased odds of diabetes. Adjusted odds ratios and credible intervals are shown in Table 4.3.

Table 4. 3: Results from multivariate logistic regression model for hypertension and diabetes by demographic, socioeconomic, anthropometric, and behavioral factors.

VARIABLE	Description	HYPERTENSION	DIABETES
		aOR (Credible Interval)	**aOR (Credible Interval)**
Demographic characteristics			
Age (years)		1.06 (1.05; 1.08)	1.05 (1.03; 1.07)
Sex			
	Male	Reference	Reference
	Female	2.45 (1.62; 3.71)	2.36 (1.41; 4.01)
Socio-economic status characteristics			
Household wealth tertile			
	1 [lowest]	Reference	Reference
	2	0.90 (0.60; 1.34)	0.59 (0.36; 0.97)
	3 [highest]	0.83 (0.51; 1.35)	0.53 (0.28; 0.98)
Support from government?			
	No	Reference	
	Yes	3.77 (0; 1.6E+8)	
Years of schooling		0.90 (0.85; 0.95)	0.93 (0.87; 0.99)
Anthropological characteristics			
BMI		1.03 (1.01; 1.05)	
Behavioral characteristics			
Add salt at table?	No		
	Yes	1.48 (0; 2.9E+8)	
Ever used tobacco?	No		Reference
	Yes		1.74 (1.04; 2.88)

aOR—adjusted odds ratio.

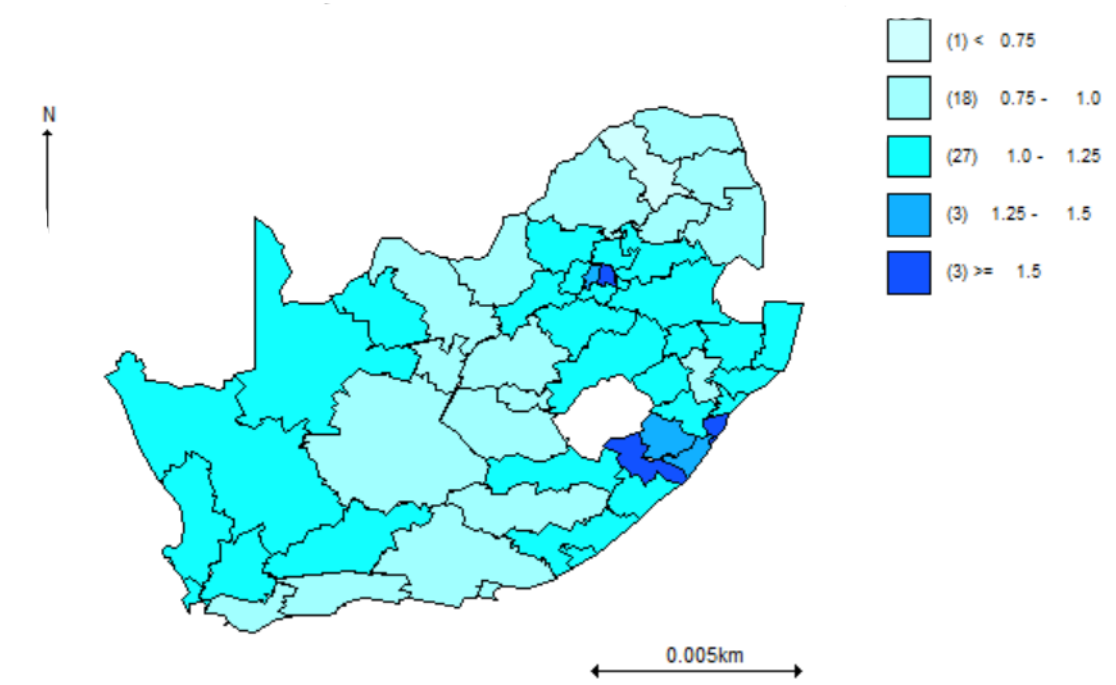
Figure 4.2 shows the disease-specific spatial distribution of the covariate-adjusted estimated odds for hypertension and diabetes. The highest spatial distribution of the covariate-adjusted estimated odds for the disease specific component of hypertension was found in some parts of Kwa-Zulu Natal, Western Cape and Gauteng areas. For diabetes, risk was found to be highest in eastern provinces of Kwa-Zulu Natal, followed by Cape Town in the Western Cape, and Mpumalanga, with the lowest in Limpopo and Eastern Cape provinces. Of note, the odds of both disease-specific components and the shared component are high in more urbanized provinces compared to the poor provinces such as the Eastern Cape, as shown in the methods. The current shared component joint spatial model had better model fit relative to a joint spatial model without the shared component (DIC = 637.2 and 1.2×10^{13}, respectively). This is in keeping with the comparison between Equation (1) versus Equations (2) and (3), which hypothesize that the joint shared model has a better model fit relative to the joint model without the shared model.



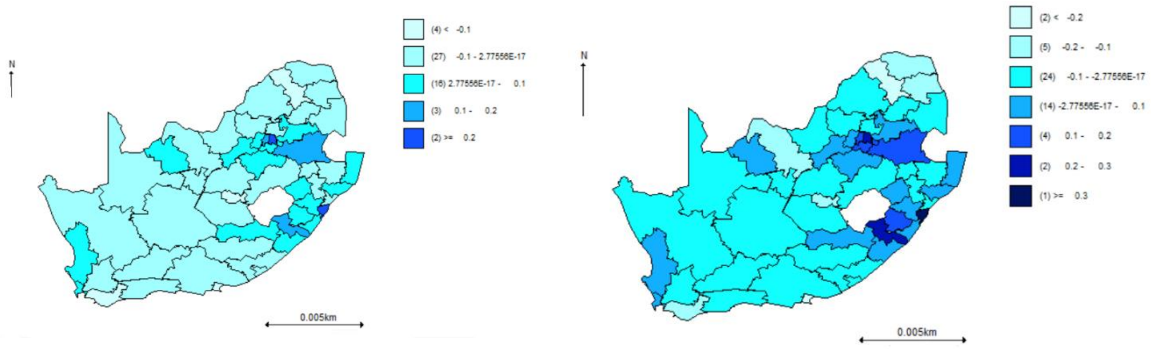
- (a) **Hypertension**
- (b) **Diabetes**

Figure 4. 2: Odds of disease-specific components for (a) hypertension, and (b) diabetes, by district, World Health Organization (WHO) Study on Global Ageing and Adult Health (SAGE) South Africa Wave 2.

Figure 4.3a displays the estimated shared component for the joint model. Of note, the shared component has distinct spatial patterns with higher values of odds in the eastern districts of Kwa-Zulu Natal and central Gauteng province. The fraction of total variation in odds ratios for hypertension and diabetes that is explained by the shared component were 0.67 and 0.53, respectively. The larger fractional contribution of hypertension to the shared component may explain the coincidental distribution between the shared estimates and hypertension estimates as shown on Figures 4.2 and 4.3a. Figure 4.3b, and c show the odds of the shared component risk contribution maps for hypertension and diabetes to indicate the absolute magnitude of shared odds for hypertension and diabetes. In addition, residual spatial effects may be more visible for diabetes than for hypertension given that the model may have explained more of hypertension.



(**a**)



(**b**) Hypertension (**c**) Diabetes

Figure 4. 3 (a): Odds of the shared component in the joint (hypertension and diabetes) shared model, (b,c): Odds of the shared component risk contribution maps for hypertension and diabetes, by district, WHO SAGE South Africa Wave 2.

4.4 Discussion

Our study used a shared joint spatial analysis to examine the spatial distribution of hypertension and diabetes and the potential role of latent and unmeasured socioeconomic status and health behavior characteristics or the social determinants of health (the shared component) on hypertension and diabetes in South Africa. Our results suggest that latent and unmeasured health behavior characteristics or the social determinants of health may have greater influence on hypertension and diabetes in the southern and central-eastern areas of the country. Common shared behavioural risks for hypertension and diabetes included in our analyses include physical activity, tobacco use, alcohol use, and salt use. The shared latent and unmeasured health behavior characteristics in our model may include ecological factors and environmental determinants such as population density, pollution, transport, power, and local food environment. The shared component specifically indicates patterns of unobserved common effects and risk factors for hypertension and diabetes in the Western Cape, Gauteng, Kwa-Zulu Natal and Mpumalanga. The spatial distribution of the shared component of hypertension and diabetes in these areas is consistent with the distribution of established risk factors such as adoption of a sedentary lifestyle and urbanization in these regions [74]. The clusters also coincide with regions with high levels of socio-economic inequality [74]. In addition, the finding that the shared component indicated a greater influence on hypertension and diabetes in the Western Cape is consistent with previous literature. Joint disease mapping models have been previously used in the investigation of the distribution of cardiovascular conditions [70,76,77]. In a study by Kandala and colleagues aimed at estimating the spatial coexistence of coronary heart disease (CHD), hypertension, hypercholesterolemia, and stroke among 13,827 South African adults from the South African Health and Demographic Survey found that the shared component, which they took to represent “nutrition and other lifestyle factors” not controlled for in their model, had a greater effect on cardiovascular disease prevalence in Western Cape and Northern Cape [74]. This is despite the fact that their study was representative of the general adult population (18+ years) while our study is focused on adult population age 50+ years [74].

The high prevalence of diabetes is indicative of the social determinants of health showing that these cardiometabolic conditions disproportionately impact vulnerable members of our society in our setting as described in the methods. A population-level effort to reduce risks for hypertension (salt intake) was instituted in South Africa starting in 2016 with an aim to contribute to lowering high blood pressure across all provinces [103]. The finding in this study that older age, higher BMI, and being female were associated with hypertension was in

agreement with previous literature in univariate analyses [104]. Unlike some other studies, this study did not find being physically inactive to be associated with hypertension and diabetes among the older adult population [127-130]. This could be due to the fact that physical activity was based on self-report rather than an objective indicator and can thus be affected by possible recall bias and social desirability bias. Previous studies conducted in Europe and North America have shown that the prevalence of hypertension is consistently higher among men compared to women across different countries [131]. This could be attributed to the fact that in our population, healthcare utilization has been found to be associated with sex, with women being more likely to seek healthcare compared to men [132,133]. In addition, high BMI has been found to be associated with increased higher risk of cardiovascular diseases, particularly hypertension [134]. Obesity has also been reported to increase cardiovascular disease risk [104]. In addition, our findings that diabetes was associated with lower education and lower household wealth category were in keeping with previous literature [135]. This could be attributed to limited access to healthcare, and poor nutrition among participants in the lower wealth tertiles. Thus, this could suggest that the social determinants of health show that hypertension and diabetes disproportionately impact vulnerable members of society in the older South African population.

A key strength of SAGE is that it consists of nationally representative samples for the older adult population, with high response rates. The current study should, however, be viewed in light of the following limitations. Firstly, our models assume that the shared and specific components are independent, which ignores the possibility of interactions between the true covariates. Secondly, the analysed health data pose a possibility of under-reporting of diabetes in addition to the lack of objective measurements on habits such as tobacco use, alcohol use and physical activity. Furthermore, SAGE data had a larger representation of the Black African population and older adults (aged 50 years and older). We were not able to explore the influence of ethnicity on our models as the current sample only had 3% white participation rate as compared to the estimated 9% representation of the white population within South Africa. Nonetheless, the shared component model used in this study may be extended to the joint analysis of three or more diseases to understand unexplained common risk factors. Furthermore, joint modelling helps to stabilize parameter estimates in small area estimation where sample sizes at sub-regions with respect to each disease are small. In epidemiology, joint modelling may be useful in identifying similar patterns of disease and understanding diseases association.

4.5 Conclusion

The co-occurrence of hypertension and diabetes remains a concern in South Africa. Our study further showed how a shared component is distributed across South Africa among the older adult population. We further illustrate how this shared component is likely to influence the geographic distribution of hypertension, and diabetes in South Africa. Policymakers may potentially use our spatial results for purposes of resource allocation and education in public health programs targeted to reduce the burden of hypertension and diabetes in South Africa, and also to manage this co-occurrence concurrently. In addition, further research using similar shared component joint models may focus on extending these models for multiple diseases with ecological factors and incorporating of sampling weights in the spatial analyses.

In chapter 5, I explore further the association between MM and depression with several socio-economic, demographic, behavioural, and environmental factors.

CHAPTER 5: UNDERSTANDING THE INTER-RELATIONSHIPS BETWEEN SOCIO-ECONOMIC, SOCIO-DEMOGRAPHIC, BEHAVIOURAL, AND ENVIRONMENTAL FACTORS FOR MULTIMORBIDITY: A STRUCTURAL EQUATION MODELLING APPROACH

Paper prepared for submission to PLOS-ONE by January 2022

Author Contributions

Conceptualization, G.C., I.M., L.J.W., P.K. and L.K.M.; methodology, G.C. and I.M.; software, STATA, OPENBUGS; formal analysis, G.C.; investigation, P.K.; resources, WHO.; data curation, WHO; writing—original draft preparation, G.C.; writing—review and editing, G.C.; visualization, G.C.; supervision, I.M., L.J.W., L.K.M.; project administration, P.K.; funding acquisition, P.K. All authors have read and agreed to the published version of the manuscript.

5.1 Introduction

South African data from the Study on global AGEing and adult health (SAGE) has shown that the prevalence of multimorbidity (MM) in South Africa was 21% among adults aged 45 years and above, 21.6% among 50–64-year-olds, and 30.1% in adults 65 years and older [31,136]. Despite such evidence on MM, traditionally NCDs are managed separately, without adequate consideration of multimorbidity in individual patients [113]. Factors such as population growth, the increasing average age of the world's population, progress made in reducing mortality rates of communicable diseases, and improvements in maternal and neonatal care, have inadvertently led to the significant increase in NCDs [137,138]. At the same time South Africa is experiencing a double burden of chronic disease as the incidences of both communicable and non-communicable disease are on the rise, with the WHO reporting the burden from NCDs in South Africa to be two to three times higher than in developed countries [13]. Most significantly the burden of human immunodeficiency virus (HIV) and tuberculosis (TB) remain high, despite many interventions [139].

Factors that are associated with NCD risk in South Africa include age, area of residence, tobacco use, physical inactivity and unhealthy diets [13]. Data from the National Income Dynamics Study (NIDS) on approximately 28 000 people investigated the association between socioeconomic disadvantage and multimorbidity in South Africa between 2008 and 2012. The study found that respondents who were socioeconomically deprived had a two-fold higher odds of having multimorbidity compared to the non-deprived [11]. In addition, studies investigating the role of specific co-occurrence of diseases have indicated that MM can particularly impact the development and the course of depression, we therefore analysed depression separately from MM [140-142].

In examining the association between different factors and MM, previous studies have used approaches such as linear and logistic regression [24,50]. We used a conceptual framework (Figure 5.1) to explore the inter-relationships between factors (socio-economic, demographic, behavioural and environmental factors) and their association with MM and depression [143]. To achieve the goal, this study utilised logistic regression and generalized structural equation modelling (gSEM). This provides a flexible and general framework to test several potential relationships between a number of variables in the model and thereby provide estimates that are more reflective of the true magnitude of the effect [79]. [95]. gSEM, as an extension of general linear model (GLM) procedures, has been preferred over other GLM methods for a number of reasons that include: (i) the ability to model all pathways simultaneously using a

set of equations (ii) the allowance for estimation simultaneous paths effects; and (iii) the ability to quantify each factor's contribution to the covariance structure [90]. To our knowledge, this will be the first study to utilise gSEM to investigate different factors associated with multimorbidity and depression among middle aged and older adults (45 years and older) in South Africa.

The current study, therefore, aims to examine the association between socio-economic, demographic, behavioural and environmental variables, and MM and depression, using logistic regression and general structural equation modelling, in middle aged and older adults using a nationally representative sample in South Africa.

5.2 Methods

Study Design and Participants

The current study used the South African Wave 2 data from the World Health Organization Study on global AGEing and adult health (WHO SAGE). The WHO SAGE study is a multi-country longitudinal study including China, Ghana, India, Mexico, and the Russian Federation, that aims to examine the health and wellbeing of nationally representative adult populations aged 50+ years in over 42,000 participants, with a comparison population aged between 18 to 49 years [9]. Further details are available on the WHO SAGE website (http://www.who.int/healthinfo/sage/en/). WHO SAGE data collection for Wave 1 was carried out between 2007 and 2010. The current study includes data collected in 2014/5 using participants (n=1,967), who were at least 45 years of age with valid (not equal to zero) post-stratification weights and complete data on hypertension, angina, arthritis, asthma, chronic lung disease (emphysema or bronchitis, chronic obstructive pulmonary disease), depression and diabetes.

Measurement of variables

Outcome Variables

Multimorbidity and depression were the primary outcome variables. Data were collected via measurement and/or self-report. Hypertension status was determined as a measured average systolic blood pressure (SBP) reading of ≥140 mmHg and/or an average diastolic blood pressure (DBP) reading of ≥90 mmHg and/or current use (within the last 2 weeks) of antihypertensive medication [104]. Participants reported whether they had ever received a medical diagnosis for angina, arthritis, asthma, chronic lung disease (emphysema or bronchitis, chronic obstructive pulmonary disease), depression and diabetes. These six self-

reported NCDs were assessed through the question “Have you ever been told by a health professional/doctor that you have [disease name]?”. For each of these six NCDs, a binary variable was created to indicate: (0) no NCD and (1) presence of an NCD. Multimorbidity was defined by the presence of 2 or more NCDs. The categories were as follows: 0) no NCD or 1 NCD and (1) presence of 2 or more NCDs.

Exposure variables

Sociodemographic data included sex and age. Socioeconomic status (SES) variables included years of schooling, household wealth and social cohesion. For household wealth, indices were derived from principal components analysis (PCA) using household assets determined by asset count, and then grouped into tertiles, as previously applied [107,106,105,136]. Social cohesion was measured with nine items, starting with the introduction ‘How often in the last 12 months have you attended any… [group, club, society, union, organizational meeting]?’ Response options ranged from never = (1) to daily = (5). The scores assigned to each of the items were ‘never’ (1), ‘once or twice a year’ (2), ‘once or twice per month’ (3), ‘once or twice per week’ (4), and ‘daily’ (5) were summed. These responses were used to create a single score for overall social cohesion [144]. Cronbach’s alpha for this social cohesion index in this sample was 0.85. The social cohesion score was then categorized into tertiles of low, medium, and high. Environmental variables included environmental safety and place of residence “urban” (reference) vs. “rural” – these were predetermined in SAGE sampling. Environmental safety was measured by asking participants how safe they felt walking alone after dark. Responses were based on a 5-point Likert scale (1) =completely safe (reference) to (5) =not safe at all. For the current study, the totals from the responses were further categorized into tertiles namely: “unsafe”, “moderate” and “safe”. Behavioural variables included ever used alcohol, ever used tobacco (smoked and smokeless), add salt to food at the table (yes/no), participation in self-reported vigorous intensity activity (yes/no – “Does your work involve vigorous-intensity activity that causes large increases in breathing or heart rate, [like heavy lifting, digging or chopping wood] for at least 10 minutes continuously?”, and “Do you do any vigorous intensity sports, fitness or recreational (leisure) activities that cause large increases in breathing or heart rate [like running or football], for at least 10 minutes continuously?”), using response items from the Global Physical Activity Questionnaire (GPAQ) [145], and self-rated sleep quality (very good/good, moderate or poor/very poor) as reported previously [9]. Anthropometric measures included weight, height, hip and waist circumference and were measured in accordance with WHO standardised techniques with

fieldwork teams trained by WHO staff. Details about the WHO standardised interview and direct measurement techniques are described elsewhere [9]. Body Mass Index (BMI; weight, kg / height, m^2), and waist to height ratio [waist (cm) / height (cm)] were calculated.

Statistical analysis

Stata statistical software, version 16.1 (StataCorp, College Station, TX: Stata Corporation) was used for all analyses. Descriptive statistics were completed for the descriptive characteristics and tested for statistically significant associations between multimorbidity categories using chi-squared tests, and Mann-Whitney tests for the continuous characteristics, given the non-normal distributions as shown by the Shapiro Wilk test. Multivariable analyses were conducted for the association of all the characteristics in Table 5.2 with multimorbidity and depression.

Multivariable analyses performed in gSEM were based and guided by an *a priori* conceptual model (Figure 5.1), and gSEM analysis was used to assess associations on all pathways.

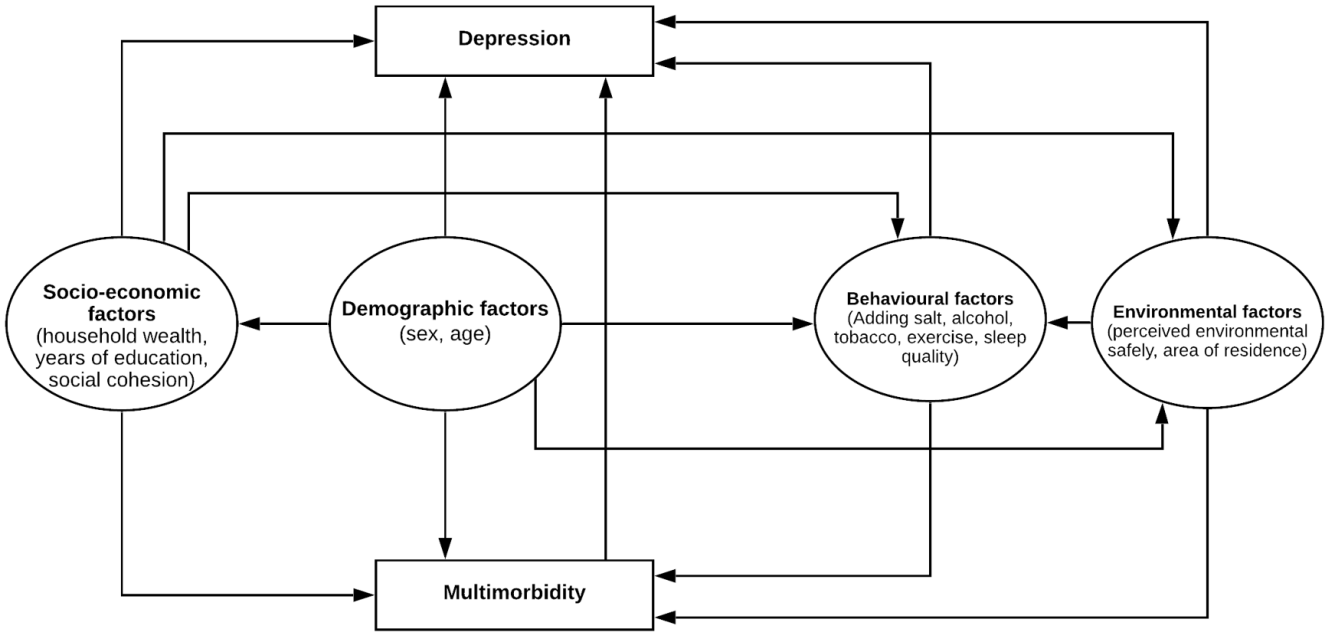


Figure 5. 1: Conceptual framework for the study

Based on literature and expert knowledge, we hypothesized relationships *a priori* between the variables that were analysed. Using this framework gSEM was applied to estimate the associations between socio-demographic, socio-economic, behavioural and environmental factors themselves, and then with multimorbidity and depression. Model estimation in the specified structural model was carried out using an iterative procedure of maximum

likelihood. We explored whether demographic, socio-economic, behavioural and environmental factors had associations with multimorbidity and depression separately. We utilized stepwise logistic regression models from *a priori* knowledge for depression and multimorbidity to select candidate variables for the gSEM using a p-value of 0.20. Variables selected are shown in Figure 5.2 in the results section.

5.3 Results

Of the 1,967 participants, 21% (n=415) had multimorbidity. Cross tabulations show that socio-demographic factors that were associated with a higher prevalence of MM were age (older) and being female. In addition, higher anthropometric measurements (BMI, waist and hip circumferences, and waist-to-height ratio) were in participants with multimorbidity. The prevalence of multimorbidity was higher among participants who perceived their environment to be unsafe, who did not participate in vigorous exercise, and who reported adding salt to their food at the table (See Table 5.1).

Table 5. 1: Prevalence of multimorbidity presented by explanatory factors.

VARIABLE	TOTAL	MULTIMORBIDITY		P-value
	(N=1,967)	No (N=1,552)	Yes (N=415)	
Demographic variables				
Age (years)	62 (54, 70)	60 (53, 69)	65 (58, 72)	<0.001
Sex				
Male	623 (31.7%)	523 (35.0%)	100 (21.2%)	<0.001
Female	1344 (68.3%)	972 (65.0%)	372 (78.8%)	
Anthropometric variables				
Waist circumference (cm)				
Median (IQR)	95 (83, 106)	94 (82, 105)	98 (87, 109)	<0.001
Hip circumference (cm)				
Median (IQR)	101 (90, 113)	100 (90, 112)	106 (93, 116)	<0.001
Waist-to-height ratio				
Median (IQR)	0.61 (0.53, 0.69)	0.60 (0.52, 0.68)	0.64 (0.56, 0.72)	<0.001
Socioeconomic variables				
Years educated				
Median (IQR)	8 (6, 11)	8 (6, 11)	8 (6, 10)	0.007
Household wealth tertile				
1 [lowest]	540 (33.4%)	393 (32.1%)	147 (37.3%)	0.026
2	539 (33.3%)	402 (32.8%)	137 (34.8%)	
3 [highest]	539 (33.3%)	429 (35.0%)	110 (27.9%)	
Behavioral variables				
Ever used alcohol				
Yes	349 (17.8%)	268 (18.0%)	81 (17.3%)	0.715
No	1608 (82.2%)	1220 (82.0%)	388 (82.7%)	
Ever used tobacco				
Yes	376 (19.2%)	281 (18.9%)	95 (20.3%)	0.498
No	1580 (80.8%)	1207 (81.1%)	373 (79.7%)	
Add salt at table				
Yes	1312 (67.0%)	1019 (68.5%)	293 (62.3%)	0.014
No	646 (33.0%)	469 (31.5%)	177 (37.7%)	
Self-reported vigorous exercise				
Yes	571 (29.3%)	411 (27.8%)	160 (34.3%)	0.007
No	1376 (70.7%)	1069 (72.2%)	307 (65.7%)	
Sleep quality				
Good	1572 (81.2%)	1237 (83.9%)	335 (72.5%)	<0.001
Bad	364 (18.8%)	237 (16.1%)	127 (27.5%)	
Environmental variables				
Residence				
Urban	1385 (70.4%)	1039 (69.5%)	346 (73.3%)	0.114
Rural	582 (29.6%)	456 (30.5%)	126 (26.7%)	
Social cohesion index				
Low	605 (33.4%)	455 (32.8%)	150 (35.3%)	0.429
Medium	605 (33.4%)	461 (33.2%)	144 (33.9%)	
High	604 (33.3%)	473 (34.1%)	131 (30.8%)	
Perceived environmental safety				
Not safe	887 (46.2%)	632 (43.4%)	255 (55.2%)	<0.001
Moderate	591 (30.8%)	470 (32.3%)	121 (26.2%)	
Safe	441 (23.0%)	355 (24.4%)	86 (18.6%)	

All data is presented as frequency and percentage, unless otherwise stated, IQR - inter-quartile range. MM – presence of 2 or more of the 6 NCDs (hypertension, angina, arthritis, asthma, chronic lung disease (emphysema or bronchitis, chronic obstructive pulmonary disease), and diabetes). P-values shown are from Chi-squared tests and Mann-Whitney U test for categorial and continuous variables, respectively.

Multivariable logistic regression models for factors associated with multimorbidity and depression are presented in Table 5.2. The factors that were associated with greater odds of having multimorbidity were being female [Adjusted odds ratio (aOR)=1.93; 95% Confidence Interval: 1.02; 3.62], feeling “unsafe” [aOR=2.07; 95% Confidence Interval: 1.25; 3.42] and older age [aOR=1.05; 95% Confidence Interval: 1.02; 1.08]. Factors that were associated with

risk of depression were being female, which was associated with higher risk of depression, and being in the wealthiest tertile was associated with lower risk of depression. Adults with multimorbidity had two-fold higher odds (and hence risk) of having depression, though this was not statistically significant [aOR=1.99; 95% Confidence Interval: 0.92; 4.30].

Table 5. 2: Results of fitting multivariable logistic regression models for factors associated with depression and multimorbidity.

Variable	Depression		Multimorbidity (excluding depression)	
	aOR (95% CI)	P-value	aOR (95% CI)	P-value
Multimorbidity				
No	Reference			
Yes	1.99 (0.92; 4.30)	0.081		
Demographic variables				
Sex				
Male	Reference		Reference	
Female	8.34 (1.83; 37.94)	0.006	1.93 (1.02; 3.62)	0.042
Age	0.96 (0.91; 1.02)	0.192	1.05 (1.02; 1.08)	0.002
Socio-economic variables				
Household wealth tertile				
1 [lowest]	Reference		Reference	
2	1.87 (0.53; 6.62)	0.329	1.44 (0.66; 3.16)	0.362
3 [highest]	0.15 (0.03; 0.80)	0.027	1.05 (0.41; 2.67)	0.923
Years of schooling	1.07 (0.93; 1.24)	0.335	0.99 (0.90; 1.08)	0.757
Social cohesion index				
Low	Reference		Reference	
Medium	1.57 (0.52; 4.70)	0.420	0.53 (0.27; 1.05)	0.069
High	0.46 (0.14; 1.58)	0.218	0.85 (0.47; 1.51)	0.574
Behavioural variables				
Alcohol				
No	Reference		Reference	
Yes	0.13 (0.01; 1.54)	0.105	1.46 (0.57; 3.76)	0.431
Tobacco				
No	Reference		Reference	
Yes	2.80 (0.82; 9.60)	0.102	1.80 (0.74; 4.41)	0.195
Add salt at table				
No	Reference		Reference	
Yes	0.56 (0.22; 1.44)	0.228	0.59 (0.31; 1.10)	0.095
Exercise				
No	Reference		Reference	
Yes	1.31 (0.39; 4.46)	0.663	1.24 (0.65; 2.33)	0.512
Sleep quality				
Good/Very good	Reference		Reference	
Moderate	2.15 (0.73; 6.35)	0.165	1.00 (0.52; 1.94)	0.998
Poor/Very poor	1.03 (0.18; 5.91)	0.976	0.62 (0.21; 1.78)	0.370
Environmental variables				
Area of residence				
Rural	Reference		Reference	
Urban	1.39 (0.59; 3.29)	0.046	1.08 (0.72; 1.62)	0.248
Perceived environmental safety				
Safe	Reference		Reference	
Moderate	2.43 (0.63; 9.37)	0.588	1.38 (0.80; 2.41)	0.250
Unsafe	2.54 (0.71; 9.16)	0.152	2.07 (1.25; 3.42)	0.005

aOR – adjusted odd ratio (95% confidence intervals in parenthesis)

Results from the gSEM for the association between demographic, socio-economic, behavioural, and environmental variables, and multimorbidity and depression are shown in Figure 5.2.

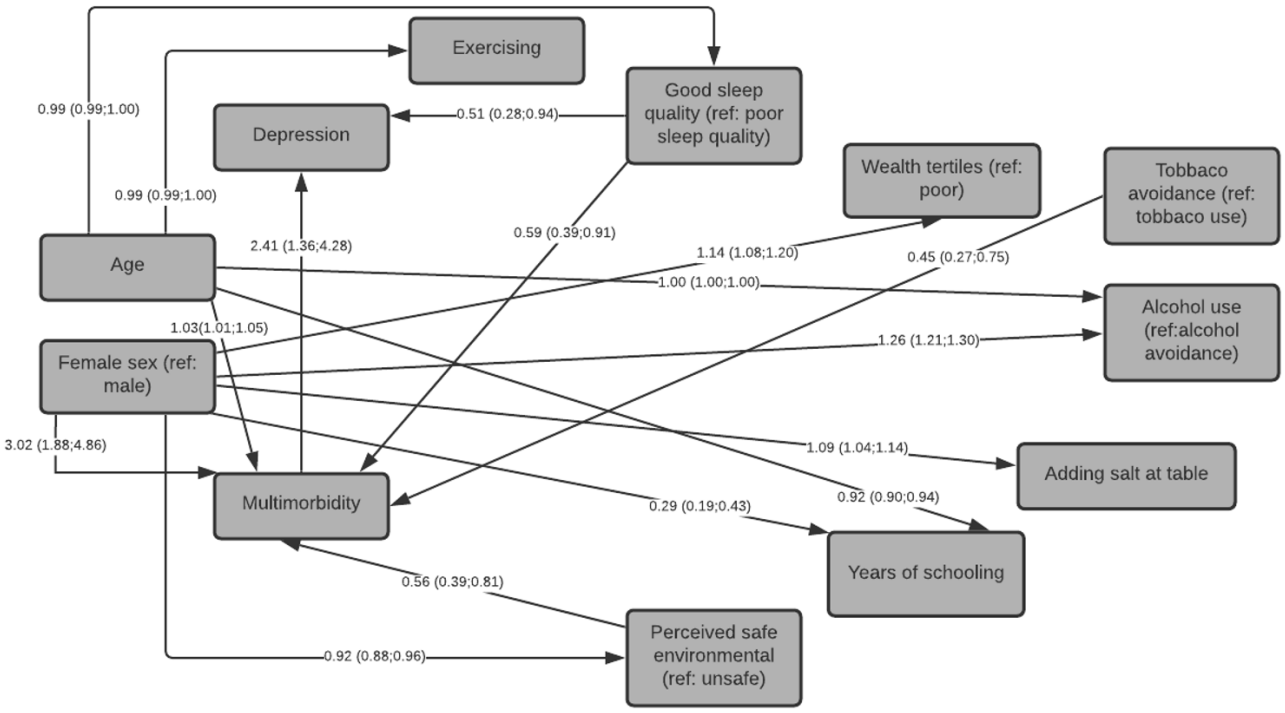


Figure 5. 2: Generalized structural equation model results for the hypothesized model.

aOR – adjusted odd ratio (95% confidence intervals in parenthesis). Only statistically significant paths presented. Sleep quality was categorized as good/poor. Tertiles were categorized as poor/rich. Tobacco and alcohol were categorized as avoidance/use. Environmental safety was categorized as safe/unsafe. Exercise and adding salt at table were categorized as yes/no. Continuous variables were age, BMI, and years of schooling.

We obtained odds ratios and 95% confidence intervals from the gSEM model to describe the relationship between the demographic, socio-economic, behavioural, and environmental characteristics, and with multimorbidity and depression, as shown in Figure 5.2 above. In this model, demographic factors (older age and being female), and behavioural factors (history of tobacco use and poor sleep quality) and perceived unsafe environment were significantly associated with multimorbidity.

5.4 Discussion

Our aim was to examine the association between socio-economic, demographic, behavioural and environmental variables, and their association with MM and depression, using logistic regression and general structural equation modelling.

Utilising logistic regression and gSEM, we showed that perceived “unsafe” environment had a direct effect on MM. Our results also confirmed the association between MM and older age, female sex, tobacco use and poor sleep quality. Of note, participants with MM were twice as likely to have depression indicating a need for closer collaborations between primary care and mental health services.

Of importance, we found that perceived “safe” environment offered 44% reduced odds of MM compared to perceived “unsafe” environment. Andrews and colleagues examined relationships between perceived/objective neighbourhood characteristics, depression, and cardio-vascular disease markers within Washington DC, USA among older adults with a mean age of 59 years. The study found that more favourable perceptions of neighbourhood physical/social environment and social cohesion were associated with decreasing depressive symptom score. Adjusting for other covariates, they found that better overall neighbourhood perceptions were related to a 0.20 unit decrease in depressive symptom scores [146].

In addition, the gSEM results showed a statistically significant 2.4-fold increased risk of depression among participants with MM compared to 1.99-fold increased risk when using logistic regression. This underscores the differences in estimates when modelled simultaneously rather than singly. However, this may be due to the direct MM effect in conjunction with indirect effects of other different factors as shown.

In keeping with previous studies in South Africa, age was associated with higher odds of multimorbidity [31,147,148]. For every year increase, the odds of MM increased by 1%. While our results show increased risk of MM in women, similar to the results of Gamma et al and Kruger et al, not all studies have found this [149,150,147,11,148,121].

We found no association between socioeconomic status and multimorbidity after adjusting for other covariates. Literature from many low and middle income countries is not conclusive as some studies has shown NCDs to be more prevalent in individuals from more affluent communities and households [85], other literature has shown socioeconomic deprivation and lower education level to be associated with multimorbidity [11,147,31]. Conversely, previous population-based studies (mostly from high income countries) indicate that better educated individuals have less risk of chronic diseases such as cardiovascular diseases, as well as multimorbidity [151,7]. Improved education could reflect a better use of information that improves access to healthcare and also indirectly reduces behavioural risk factors [7].

Elements of social organization, such as interpersonal trust, reciprocity norms, and engagements with community and neighbourhood, known as social capital, have been shown to be beneficial for health [152]. In our study, social cohesion was not associated with multimorbidity. Previous literature has shown that though social capital (measured by network membership) plays a significant role in an individual's well-being [152], its relationship with health outcomes varies from being protective in nature [153] to lack of an association [154] and in some instances destructive [155]. However, the inconsistency in the nature of association between social capital and health outcomes observed across literature may be due to variations in definitions and measurements [156]. It should also be noted that the measure of social capital used in this study did not include family ties, which have been shown to be important in managing chronic disease [18]. Future research should test for the effect of family ties, as this could have an impact on the occurrence of multimorbidity.

In keeping with previous studies, our study showed an association between depression and multimorbidity. Specifically, studies have identified the effects of multimorbidity on suffering from depression [157]. Several studies have shown depression to be associated with cardiovascular disease, particularly concurrent hypertension and to predict incident cardiovascular disease in the future [158-160].This suggests that in the management of NCDs, depression should be considered. Efforts may include depression screening in people with other NCDs. Furthermore, the odds of depression were found to decrease with increasing socioeconomic status, though it was not statistically significant. In a National Population Health Survey in Canada, Wang and colleagues also suggest that financial burden in people of lower socioeconomic status placed people at risk of depression [161]. In addition, in keeping with previous literature, we found tobacco users to have a higher probability of MM compared to non-tobacco users[114-116]. For example, in a study by Fonda and colleagues aimed at examining the clustering of post-traumatic stress disorder, depressive disorders, and clinically significant pain among 433 deployed veterans in Boston (USA) aged 18 to 65 years, tobacco smokers had 3.5 increased likelihood for multimorbidity [117].

The current study has the following strengths; Firstly, this is the first comprehensive study on factors associated with the MM and depression simultaneously in the gSEM in a low-middle-income country. Secondly, a key strength of the Study on global AGEing and adult health (SAGE) is that it consists of nationally representative samples for participants aged 50 and above, with high response rates. However, the findings from this study should be viewed

considering some limitations. Firstly, since the current study design is cross-sectional in nature, we could not determine the temporal sequence or causality. Second, data on most of the chronic diseases, and all behavioural variables (including tobacco use), was based on self-report, and can thus be affected by possible recall bias and social desirability bias. In addition, the definitions of alcohol use and tobacco use in our study were broad and do not capture the quantities and frequency of consumption, potentially explaining the lack of association found. Furthermore, there is no statistical test to objectively test goodness of fit for gSEM available in Stata/IC and gSEM requires a simple model in model specification to ensure model convergence. Finally, the chronic diseases included in this analysis were limited to those included in the SAGE study and only to NCDs. This may miss other conditions present in this population, such as dementia, cancers, and chronic infectious diseases (HIV) and therefore may have resulted in an underestimation of multimorbidity prevalence.

In conclusion, these findings guide researchers to consider utilizing gSEM in analysing future studies collecting MM data. Given that the prevalence of MM remains high, the primary health care (PHC) system in South Africa should focus on capacity building among primary healthcare workers in the management of multiple conditions. In addition, our findings suggest policy makers need to prioritize the enhancement of environmental safety among the older population and focus on females and tobacco users in prevention programs.

Chapter 6 gives a summative narration of the thesis, from the aim and objectives through the conclusions.

CHAPTER 6: DISCUSSION AND CONCLUSIONS

This concluding chapter gives an overview of the aim of the thesis, summary and implications of the findings, limitations and strengths of the study, recommendations for future research and conclusions.

6.1 Overview of research question and aim of the thesis

This thesis provides evidence of the prevalence and risk factors for MM, and on the spatial patterns of common NCDs that co-occur in middle aged and older adult South Africans from the WHO SAGE cohort. This aim was addressed in three parts (objectives):

1. To determine the prevalence of multimorbidity in a cohort of South African adults over the age of 50 years, to determine the co-occurrence of chronic diseases, and to identify the demographic, anthropometric and behavioural factors associated with three different multimorbidity classes.
2. To examine the spatial distribution of hypertension and diabetes, and the distribution of shared unmeasured characteristics on hypertension and diabetes, in South African middle aged and older adults.
3. To determine the complex inter-relationships between socio-economic, demographic, behavioural, and environmental factors that are associated with multimorbidity and depression in South African middle aged and older adults.

Table 6.2 Summary of objectives and key findings

OBJECTIVE	CHAPTER	KEY FINDINGS
1. To determine the prevalence of multimorbidity in a cohort of South African adults over the age of 45 years, to determine the co-occurrence of chronic diseases, and to identify the demographic, anthropometric and behavioural factors associated with the different multimorbidity classes.	3	• The prevalence of MM • The latent class analysis the age of 45 years into *concordant MM* (11%) • DESCRIBE THE THRE • When compared to the *d* were associated with be tobacco use, and older a *discordant MM* group.
2. To examine the spatial distribution of hypertension and diabetes, and the distribution of shared unmeasured characteristics on hypertension and diabetes, in South African middle aged and older adults.	4	• The shared component patterns with higher odd central Gauteng provinc • The larger fractional co component may explain estimates and hypertens • The odds of the shared hypertension and diabet odds for hypertension a may be more visible for model may have explai
3. To determine the complex inter-relationships between socio-economic, demographic, behavioural, and environmental factors that are associated with multimorbidity and depression in South African middle aged and older adults.	5	• There were direct assoc environments, older age sleep quality are associa • In addition, the model s aforementioned factors between age and sex wi • gSEM results showed a with MM compared to 1

6.3 Summary and implications of study findings

The findings of this thesis suggest that one in five adults aged 45 and above in South Africa experience and live with multimorbidity, with a high rate of co-occurrence of hypertension and diabetes. Moreover, this thesis provides evidence in South Africa for an association between multimorbidity and depression, suggesting the need for patients with multimorbidity to be screened for depression A better understanding of the pathway between multimorbidity and depression is critical to understand not only the aging trajectories of individuals with multimorbidity and depression, but also to potentially develop effective interventions aimed at limiting the burden of these two highly debilitating conditions. . Also, there is a need to focus on targeted interventions for MM in those who have been identified to be at higher risk including females, middle aged to older adults, as well as people in perceived "unsafe" areas and tobacco users.

Multimorbidity presents multiple challenges for primary health care providers [21,162,22]. The management of NCDs in South Africa, during a period where the public health sector is dominated by HIV/AIDS, TB, and COVID-19, has met with many challenges resulting in delayed management of conditions. Furthermore, the public health focus on the management of NCDs is disease focused and places emphasis on vertical implementation that inadequately deals with comorbid conditions. To provide better care for individuals with comorbid conditions, South Africa implemented the integrated chronic disease management (ICDM) plan in 2014 for primary health care [113]. This model focuses on a systems perspective and addresses primary prevention, (which includes early detection, appropriate screening, and surveillance), secondary prevention and tertiary prevention. The main aim of ICDM is to ensure early detection and appropriate management of high-risks patients. However, implementation of this model has not been particularly successful and many programmes are still disease focused and still practice vertical implementation which does not consider comorbidities [84]. In addition, scale-up and sustainability of the approach have proven a problem [83].

The SAGE data used for the current thesis were collected prior to the current COVID-19 pandemic. However, this pandemic has presented more difficulties in the management of NCD care. Diagnostic services, physician consultation, transport arrangements, financial constraints, mandatory self-isolation, the need for social distancing and fear of visiting hospitals for risk of COVID-19 infections have presented challenges with reduced uptake of NCD services [163,164]. A cohesive doctor-pharmacy-patient engagement is vital for managing NCD care during a pandemic. During emergencies, changes in dispensing

practices such as duration and person, and service provision closer to the patient are crucial. Additionally, community NCD care-related health literacy and community health models, including home-based NCD treatment, should be encouraged.

Furthermore, we defined co-occurrence as geographical bounding of hypertension and diabetes, where cases are closer to other cases than cases are to non-cases, and this is unlikely to have occurred by chance [165]. Clusters (districts) where the multimorbidity is more pronounced are known as “hotspots”. Understanding the spatial distribution of clustering may potentially guide healthcare policy reform in managing multi-morbidities in the South African healthcare system and explore unmeasured shared risks for multimorbidity. We established that latent and unmeasured health behaviour characteristics which may represent social determinants of health may have greater influence on hypertension and diabetes in the southern and central-eastern areas of the country. Common shared behavioural risks for hypertension and diabetes included in our analyses were insufficient physical activity, tobacco use, alcohol use, and salt use. Additionally, the shared latent and unmeasured health structural and environmental characteristics in our model may be pollution, sanitation, water, transport, and local food environment. These environmental risk factors cluster in groups of individuals as a direct result of their socioeconomic status and their living conditions [166]. In addition, the local food environment in South Africa has been affected by globalisation and the nutrition transition taking place. This nutrition transition has resulted in many rural populations purchasing lower priced processed foods, as the healthier options are usually higher priced and out of reach for most South Africans, thereby placing them at a higher risk of diabetes and hypertension, as shown by increased weight among half of South Africans due to unhealthy foods [167-170].

This observed early detection of risk factors points to the fact that to be effective, MM prevention programs/ intervention strategies must not only be targeted towards the elderly but start from young adulthood. To curb the rise of hypertension, a population level effort to reduce risks for hypertension (salt intake) was instituted in South Africa starting in 2016 with an aim to contribute to lowering high blood pressure across all provinces [171]. Urbanisation, altering demographics (aging), changing social behaviours and suboptimal public health facilities require deeper investigation within South Africa as a unique example of socio-economic shifts and continued urbanization [143].

6.4 Limitations and strengths

Limitations and strengths are outlined in detail in each of the three results chapters. A key strength of SAGE is that it consists of a large, nationally representative sample of the middle aged and older adult population (over 50 years), with high response rates. In addition, applying unique statistical techniques such as explorative LCA makes no *a priori* assumptions about the number of latent classes and estimates the number of classes starting with a two-class model and increasing the number of latent classes in a stepwise fashion. Key limitations of the thesis are summarised below.

Measurement

(i) Since the current study design is cross-sectional in nature, we could not determine the temporal sequence or causality.

(ii) Data on most of the chronic diseases, and many behavioural variables (including tobacco use), was based on self-report, and can thus be affected by possible recall bias or social desirability bias.

(iii) The definitions of alcohol use and tobacco use in our study were broad and do not capture the quantities and frequency of consumption, potentially explaining the lack of association found.

(iv) The number of diseases included in this analysis was limited to those included in the SAGE study. This may miss other conditions present in this population, such as HIV, dementia or cancers, and therefore have resulted in an underestimation of multimorbidity prevalence.

Statistical

(i) The LCA naturally combines participants without NCDs with those with mostly one NCD in the minimal MM risk group, thereby limiting the use of participants with no MM as the reference group.

(ii) The LCA procedure was explorative in nature. As such, when different criteria to determine the classes are used, researchers may argue in favour of different numbers of classes, making the procedure less objective.

(iii) There are no packages available to impute LCA class membership in order to perform a sensitivity analysis on the missing data.

(iv) There is no statistical test to objectively test goodness of fit for gSEM available in Stata/IC.

(v) The gSEM model could not produce cyclical estimates for bi-directional

associations, thereby, limiting the estimates to only one direction of association.

(vi) The current analyses using LCA, gSEM and bivariate joint spatial model were limited to complete-case analyses of data. Efforts were made to assess missingness mechanisms and there was no evidence to suspect data was missing not at random or missing at random. Complete case analysis retained a reasonably large sample (n>1000) across all models hence there was minimal loss of efficiency due to drop in sample size. On the other hand, the complete-case analyses approach may not have affected the inferences though some null bias could be possible. Given the above, either the multiple imputation or direct maximum likelihood methods in this study were not utilized. Moreover, the mainstream statistical software with relevant modelling frameworks for my study do not yet have or support these methods for dealing with missing data.

6.5 Future research and recommendations

Firstly, future research may need an analysis longitudinally (with at least 3 time points). In addition, spatio-temporal analysis may be useful to infer causality and temporal sequence between MM, depression and covariates using data at different time points. In determining co-occurrence of NCDs, the use of multilevel Bayesian networks in longitudinal studies may help in predicting the classes over time, for example, the likelihood of participants to move from *concordant MM group* to *discordant MM group* over time. In addition, spatial analyses may incorporate study survey weights.

In addition, given that South Africa and the Sub-Saharan region still have a large burden of infectious disease, MM that incorporates infectious diseases, such as TB, HIV, and COVID-19, may help in giving more insights and may guide how the chronic conditions are managed in primary healthcare services. In addition, the current objectives of this thesis should be further aligned to the United Nation Sustainable Developments Goal 3 (SDG3): Good health and wellbeing.

Furthermore, measuring NCDs more accurately, such as depression and diabetes, using continuous measures may be more insightful compared to categorized data in order to avoid loss of information rather than relying on self-report.

Lastly, studies may be conducted to assess health financing that minimizes costs by using

existing infrastructure. For example, assessing the cost effectiveness of integrating NCD care among HIV and TB compared to a vertical approach.

6.6 Conclusions

To effectively reduce diabetes, hypertension, and depression morbidity and mortality their treatment needs to be viewed holistically and within the context of each other. There is therefore an urgent need to integrate chronic disease care, including mental health, and to equip primary health care providers to manage chronic conditions together, being guided by these findings on what diseases typically cluster [172]. For example, the development of hypertension screening protocols for diabetes management and vice versa. In addition, policymakers may potentially use our spatial results for purposes of resource allocation and education in public health programs targeted to reduce the burden of hypertension and diabetes in South Africa, and to manage this co-occurrence concurrently. There is therefore a need to further explore the shared components to intervene and prevent MM onset in the respective districts. In addition, our results indicate that multimorbidity affects 1 in 5 middle aged and older adults. In our sample, risk factors for multimorbidity latent classes include age, sex, and tobacco use. Future efforts should focus on the inclusion of all frequently occurring common conditions, including infectious diseases, to evaluate clustering patterns and inform policy makers to prioritize the older population, females and tobacco users in prevention and treatment programs. However, the South African public health system is over-burdened with patients in public health facilities experiencing long queues, fewer screenings, and drug stock-outs. The results from our study may be used to screen those at risk and the spatial identification of the occurrence of hypertension and diabetes aids in tailoring potential district-level interventions.

An integrated approach is needed to prevent, screen, and treat multimorbidity and depression, as our results show that it is prevalent. An integrated approach will not only ensure that a patient comes once but will also reduce waiting times as people will visit the clinic less often. As there is a documented association between depression and MM our research therefore suggests that the screening for depression be done on all MM patients.

REFERENCES

1. Organization WH (2013) Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization,

2. Organisation WH (2011) WHO. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level: Report by the Secretariat.

3. Ahmad Kiadaliri A (2016) Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 388 (10053):1603-1658

4. Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, Casey DC, Charlson FJ, Chen AZ, Coates MM (2016) Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The lancet 388 (10053):1459-1544

5. Alwan A (2011) Global status report on noncommunicable diseases 2010. World Health Organization,

6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B (2012) Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 380 (9836):37-43. doi:10.1016/s0140-6736(12)60240-2

7. Marengoni A, Winblad B, Karp A, Fratiglioni L (2008) Prevalence of chronic diseases and multimorbidity among the elderly population in Sweden. American journal of public health 98 (7):1198-1200

8. Fortin M, Hudon C, Haggerty J, van den Akker M, Almirall J (2010) Prevalence estimates of multimorbidity: a comparative study of two sources. BMC health services research 10 (1):1-6

9. Kowal P, Chatterji S, Naidoo N, Biritwum R, Fan W, Lopez Ridaura R, Maximova T, Arokiasamy P, Phaswana-Mafuya N, Williams S (2012) Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). International Journal of Epidemiology 41 (6):1639-1649

10. Lalkhen H, Mash R (2015) Multimorbidity in non-communicable diseases in South African primary healthcare, vol 105. 2015, vol 2.

11. Weimann A, Dai D, Oni T (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. Social Science & Medicine 163:144-156

12. Nojilana B, Bradshaw D, Pillay-van Wyk V, Msemburi W, Laubscher R, Somdyala NI,

Joubert JD, Groenewald P, Dorrington RE (2016) Emerging trends in non-communicable disease mortality in South Africa, 1997-2010. South African medical journal 106 (5):477-484
13. Mayosi BM, Flisher AJ, Lalloo UG, Sitas F, Tollman SM, Bradshaw D (2009) The burden of non-communicable diseases in South Africa. Lancet (London, England) 374 (9693):934-947. doi:10.1016/s0140-6736(09)61087-4
14. National Department of Health SSA, South African Medical Research Council, ICF (2019) South Africa demographic and health survey 2016. vol 0001. Pretoria
15. Sciences AoM (2018) Multimorbidity: a priority for global health research. Academy of Medical Sciences London,
16. Folb N, Timmerman V, Levitt NS, Steyn K, Bachmann MO, Lund C, Bateman ED, Lombard C, Gaziano TA, Zwarenstein M (2015) Multimorbidity, control and treatment of noncommunicable diseases among primary healthcare attenders in the Western Cape, South Africa. South African Medical Journal 105 (8):642-647
17. Garin N, Koyanagi A, Chatterji S, Tyrovolas S, Olaya B, Leonardi M, Lara E, Koskinen S, Tobiasz-Adamczyk B, Ayuso-Mateos JL, Haro JM (2015) Global Multimorbidity Patterns: A Cross-Sectional, Population-Based, Multi-Country Study. The Journals of Gerontology: Series A 71 (2):205-214. doi:10.1093/gerona/glv128
18. Taylor AW, Price K, Gill TK, Adams R, Pilkington R, Carrangis N, Shi Z, Wilson D (2010) Multimorbidity-not just an older person's issue. Results from an Australian biomedical study. BMC public health 10 (1):1-10
19. Britt HC, Harrison CM, Miller GC, Knox SA (2008) Prevalence and patterns of multimorbidity in Australia. Medical Journal of Australia 189 (2):72-77
20. Akner G (2009) Analysis of multimorbidity in individual elderly nursing home residents. Development of a multimorbidity matrix. Archives of gerontology and geriatrics 49 (3):413-419
21. Mercer SW, Gunn J, Bower P, Wyke S, Guthrie B (2012) Managing patients with mental and physical multimorbidity. British Medical Journal Publishing Group,
22. Barnett K, Mercer S, Norbury M, Watt G, Wyke S, Guthrie B (2012) The epidemiology of multimorbidity in a large cross-sectional dataset: implications for health care, research and medical education. Lancet 380 (9836):37-43
23. Violan C, Foguet-Boreu Q, Flores-Mateo G, Salisbury C, Blom J, Freitag M, Glynn L, Muth C, Valderas JM (2014) Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. PloS one 9 (7):e102149
24. Alaba O, Chola L (2013) The social determinants of multimorbidity in South Africa. International journal for equity in health 12 (1):63

25. Pati S, Agrawal S, Swain S, Lee JT, Vellakkal S, Hussain MA, Millett C (2014) Non communicable disease multimorbidity and associated health care utilization and expenditures in India: cross-sectional study. BMC health services research 14 (1):1-9
26. Lee JT, Hamid F, Pati S, Atun R, Millett C (2015) Impact of noncommunicable disease multimorbidity on healthcare utilisation and out-of-pocket expenditures in middle-income countries: cross sectional analysis. PloS one 10 (7):e0127199
27. Eyowas FA, Schneider M, Yirdaw BA, Getahun FA (2019) Multimorbidity of chronic non-communicable diseases and its models of care in low-and middle-income countries: a scoping review protocol. BMJ open 9 (10):e033320
28. Wang SB, D'Arcy C, Yu YQ, Li B, Liu YW, Tao YC, Wu YH, Zhang QQ, Xu ZQ, Fu YL, Kou CG (2015) Prevalence and patterns of multimorbidity in northeastern China: a cross-sectional study. Public Health 129 (11):1539-1546. doi:10.1016/j.puhe.2015.06.013
29. Nunes BP, Batista SRR, Andrade FBd, Souza PRBd, Lima-Costa MF, Facchini LA (2018) Multimorbidity: the Brazilian longitudinal study of aging (ELSI-Brazil). Revista de saude publica 52
30. Nimako BA, Baiden F, Sackey SO, Binka F (2013) Multimorbidity of chronic diseases among adult patients presenting to an inner-city clinic in Ghana. Global Health 9:61. doi:10.1186/1744-8603-9-61
31. Afshar S, Roderick PJ, Kowal P, Dimitrov BD, Hill AG (2015) Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. BMC Public Health 15 (1):776
32. Roomaney RA, van Wyk B, Turawa EB, Pillay-van Wyk V (2021) Multimorbidity in South Africa: a systematic review of prevalence studies. BMJ Open 11 (10):e048676. doi:10.1136/bmjopen-2021-048676
33. Fortin M, Stewart M, Poitras M-E, Almirall J, Maddocks H (2012) A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. The Annals of Family Medicine 10 (2):142-151
34. Ofori-Asenso R, Chin KL, Curtis AJ, Zomer E, Zoungas S, Liew D (2019) Recent patterns of multimorbidity among older adults in high-income countries. Population health management 22 (2):127-137
35. Salive ME (2013) Multimorbidity in older adults. Epidemiol Rev 35:75-83. doi:10.1093/epirev/mxs009
36. Nguyen H, Manolova G, Daskalopoulou C, Vitoratou S, Prince M, Prina AM (2019) Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. Journal of comorbidity 9:2235042X19870934

37. Xu X, Mishra GD, Jones M (2017) Mapping the global research landscape and knowledge gaps on multimorbidity: a bibliometric study. Journal of global health 7 (1)
38. Roomaney RA, van Wyk B, Turawa EB, Pillay-van Wyk V (2020) Prevalence of multimorbidity in South Africa: a systematic review protocol. BMJ open 10 (12):e042889
39. Organization WH (2010) World health statistics 2010. World Health Organization,
40. Schouw D, Mash R, Kolbe-Alexander T (2018) Transforming the workplace environment to prevent non-communicable chronic diseases: participatory action research in a South African power plant. Glob Health Action 11 (1):1544336. doi:10.1080/16549716.2018.1544336
41. Ndinda C, Chilwane D, Mokomane Z (2013) Civil society activism in accessing healthcare in South Africa: technical report.
42. Bosire EN (2021) Patients' Experiences of Comorbid HIV/AIDS and Diabetes Care and Management in Soweto, South Africa. Qualitative Health Research 31 (2):373-384
43. Glynn LG, Valderas JM, Healy P, Burke E, Newell J, Gillespie P, Murphy AW (2011) The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Family practice 28 (5):516-523
44. Roberts K, Rao D, Bennett T, Loukine L, Jayaraman G (2015) Prevalence and patterns of chronic disease multimorbidity and associated determinants in Canada. Health promotion and chronic disease prevention in Canada: research, policy and practice 35 (6):87
45. Ryan BL, Allen B, Zwarenstein M, Stewart M, Glazier RH, Fortin M, Wetmore SJ, Shariff SZ (2020) Multimorbidity and mortality in Ontario, Canada: A population-based retrospective cohort study. J Comorb 10:2235042x20950598. doi:10.1177/2235042x20950598
46. Mercer S, Salisbury C, Fortin M (2014) ABC of Multimorbidity. John Wiley & Sons,
47. Pearson TA (2003) Education and income: double-edged swords in the epidemiologic transition of cardiovascular disease. Ethnicity and disease 13 (2; SUPP/2):S2-158
48. Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA (2011) Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract 61 (582):e12-e21
49. Gordon T, Booysen F, Mbonigaba J (2020) Socio-economic inequalities in the multiple dimensions of access to healthcare: the case of South Africa. BMC Public Health 20 (1):289. doi:10.1186/s12889-020-8368-7
50. Ataguba JE-O (2013) Inequalities in multimorbidity in South Africa. International journal for equity in health 12 (1):64
51. Mounce LT, Campbell JL, Henley WE, Arreal MCT, Porter I, Valderas JM (2018)

Predicting incident multimorbidity. The Annals of Family Medicine 16 (4):322-329
52. Keizer I, Gex-Fabry M, Eytan A, Bertschy G (2009) Smoking in psychiatric inpatients: Association with working status, diagnosis, comorbid substance abuse and history of suicide attempts. Addictive behaviors 34 (10):815-820
53. Nagel G, Peter R, Braig S, Hermann S, Rohrmann S, Linseisen J (2008) The impact of education on risk factors and the occurrence of multimorbidity in the EPIC-Heidelberg cohort. BMC public health 8 (1):1-10
54. Bornovalova MA, Levy R, Gratz KL, Lejuez C (2010) Understanding the heterogeneity of BPD symptoms through latent class analysis: initial results and clinical correlates among inner-city substance users. Psychological assessment 22 (2):233
55. Campbell SB, Morgan-Lopez AA, Cox MJ, McLoyd VC (2009) A latent class analysis of maternal depressive symptoms over 12 years and offspring adjustment in adolescence. Journal of abnormal psychology 118 (3):479
56. Feingold A, Tiberio SS, Capaldi DM (2014) New approaches for examining associations with latent categorical variables: applications to substance abuse and aggression. Psychology of Addictive Behaviors 28 (1):257
57. Garrett ES, Eaton WW, Zeger S (2002) Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: a latent class model approach. Statistics in medicine 21 (9):1289-1307
58. Mannarini S, Boffo M (2015) Anxiety, bulimia, drug and alcohol addiction, depression, and schizophrenia: what do you think about their aetiology, dangerousness, social distance, and treatment? A latent class analysis approach. Social Psychiatry and Psychiatric Epidemiology 50 (1):27-37
59. Romano E, Baillargeon RH, Wu H-X, Zoccolillo M, Vitaro F, Tremblay RE (2004) A new look at inter-informant agreement on conduct disorder using a latent class approach. Psychiatry research 129 (1):75-89
60. Yuan C, Wei C, Wang J, Qian H, Ye X, Liu Y, Hinds PS (2014) Self-efficacy difference among patients with cancer with different socioeconomic status: Application of latent class analysis and standardization and decomposition analysis. Cancer epidemiology 38 (3):298-306
61. Vanheule S, Desmet M, Verhaeghe P, Bogaerts S (2007) Alexithymic depression: evidence for a depression subtype? Psychotherapy and Psychosomatics 76 (5):315
62. Alkan Härtwig E, Crayen C, Heuser I, Eid M (2014) It's in the mix: psychological distress differs between combinations of alexithymic facets. Frontiers in psychology 5:1259
63. Lazarsfeld PF (1950) The logical and mathematical foundation of latent structure analysis.

Studies in Social Psychology in World War II Vol IV: Measurement and Prediction:362-412
64. Vermunt JK, Magidson J (2002) Latent class cluster analysis. Applied latent class analysis 11:89-106
65. Goodman LA (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 61 (2):215-231
66. Land KC (2001) Introduction to the special issue on finite mixture models. Sage Publications,
67. Schreiber JB, Pekarik AJ (2014) Using Latent Class Analysis versus K-means or Hierarchical Clustering to understand museum visitors. Curator: The Museum Journal 57 (1):45-59
68. Nylund KL, Asparouhov T, Muthén BO (2007) Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural equation modeling: A multidisciplinary Journal 14 (4):535-569
69. Blanton H, Jaccard J (2006) Arbitrary metrics in psychology. American Psychologist 61 (1):27
70. Dabney AR, Wakefield JC (2005) Issues in the mapping of two diseases. Statistical Methods in Medical Research 14 (1):83-112
71. Downing A, Forman D, Gilthorpe MS, Edwards KL, Manda SO (2008) Joint disease mapping using six cancers in the Yorkshire region of England. International Journal of Health Geographics 7 (1):41
72. Held L, Natário I, Fenton SE, Rue H, Becker N (2005) Towards joint disease mapping. Statistical Methods in Medical Research 14 (1):61-82
73. Kazembe LN, Muula AS, Appleton CC, Kleinschmidt I (2007) Modelling the effect of malaria endemicity on spatial variations in childhood fever, diarrhoea and pneumonia in Malawi. International Journal of Health Geographics 6 (1):33
74. Kandala N-B, Manda SO, Tigbe WW, Mwambi H, Stranges S (2014) Geographic distribution of cardiovascular comorbidities in South Africa: a national cross-sectional analysis. Journal of Applied Statistics 41 (6):1203-1216
75. Manda SO, Feltbower RG, Gilthorpe MS (2009) Investigating spatio-temporal similarities in the epidemiology of childhood leukaemia and diabetes. European Journal of Epidemiology 24 (12):743
76. Knorr-Held L, Best NG (2001) A shared component model for detecting joint and selective clustering of two diseases. Journal of the Royal Statistical Society: Series A (Statistics in Society) 164 (1):73-85
77. Langford IH, Leyland AH, Rasbash J, Goldstein H (1999) Multilevel modelling of the

geographical distributions of diseases. Journal of the Royal Statistical Society: Series C (Applied Statistics) 48 (2):253-268
78. Manda S, Feltbower R, Gilthorpe M (2012) Review and empirical comparison of joint mapping of multiple diseases. Southern African Journal of Epidemiology and Infection 27 (4):169-182
79. Alexandrowicz RW, Jahn R, Friedrich F, Unger A (2016) The importance of statistical modelling in clinical research : Comparing multidimensional Rasch-, structural equation and linear regression models for analyzing the depression of relatives of psychiatric patients. Neuropsychiatrie : Klinik, Diagnostik, Therapie und Rehabilitation : Organ der Gesellschaft Osterreichischer Nervenarzte und Psychiater 30 (2):92-102. doi:10.1007/s40211-016-0180-3
80. Kupek E (2006) Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. BMC Medical Research Methodology 6 (1):1-10. doi:10.1186/1471-2288-6-13
81. Rabe-Hesketh S, Skrondal A, Pickles A (2004) Generalized multilevel structural equation modeling. Psychometrika 69 (2):167-190. doi:10.1007/bf02295939
82. van Rensburg HC (2014) South Africa’s protracted struggle for equal distribution and equitable access–still not there. Human resources for health 12 (1):1-16
83. Mahomed OH, Asmall S (2017) Professional nurses' perceptions and experiences with the implementation of an integrated chronic care model at primary healthcare clinics in South Africa. Curationis 40:1-6
84. Mahomed OH, Asmall S, Voce A (2016) Sustainability of the integrated chronic disease management model at primary care clinics in South Africa. Afr J Prim Health Care Fam Med 8 (1):e1-e7. doi:10.4102/phcfm.v8i1.1248
85. Steyn K, Damasceno A (2006) Lifestyle and related risk factors for chronic diseases. Disease and mortality in sub-Saharan Africa 2:247-265
86. Rayner B (2010) Hypertension: detection and management in South Africa. Nephron Clinical Practice 116 (4):c269-c273
87. Haberman SJ (1979) Analysis of qualitative data. vol. 2, new developments. Academic P.,
88. Hagenaars JA (1990) Categorical longitudinal data: Log-linear panel, trend, and cohort analysis. SAGE Publications, Incorporated,
89. Vermunt JK (1997) Log-linear models for event histories, advanced quantitative techniques in the social sciences series. Thousand Oaks, CA: SAGE Publications, Inc,
90. Lei P-W, Wu Q (2007) Introduction to Structural Equation Modeling: Issues and Practical Considerations. Educational Measurement: Issues and Practice 26 (3):33-43. doi:10.1111/j.1745-3992.2007.00099.x

91. Beran TN, Violato C (2010) Structural equation modeling in medical research: a primer. BMC Research Notes 3 (1):1-10. doi:10.1186/1756-0500-3-267
92. H HR (1995) Structural Equation Modeling: Concepts, Issues and Applications. Thousand Oaks, Calif: Sage Publications
93. Hancock GR, Mueller R (2001) Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), Structural Equation Modeling: Present and Future -- Festschrift in honor of Karl Jöreskog. Lincolnwood, IL: Scientific Software International, Inc:195-216
94. Byrne BM (1994) Structural equation modeling with EQS and EQS/Windows, basic concepts, applications, and programming. Thousand Oaks London New Delhi: SAGE Publications
95. Hox JJ (2007) An Antroduction to Structural Equation Modeling. Family Science Review (11):354-373
96. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, Blumenthal D (2007) Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. Journal of general internal medicine 22 (3):391-395
97. Garin N, Koyanagi A, Olaya B, Lara E, Haro JM, Tyrovolas S, Ayuso-Mateos JL, Chatterji S, Leonardi M, Koskinen S, Tobiasz-Adamczyk B (2015) Global Multimorbidity Patterns: A Cross-Sectional, Population-Based, Multi-Country Study. The Journals of Gerontology: Series A 71 (2):205-214. doi:10.1093/gerona/glv128
98. Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L (2005) Prevalence of multimorbidity among adults seen in family practice. The Annals of Family Medicine 3 (3):223-228
99. Holden L, Scuffham PA, Hilton MF, Muspratt A, Ng S-K, Whiteford HA (2011) Patterns of multimorbidity in working Australians. Population health metrics 9 (1):15
100. Ayeni OA, Norris SA, Joffe M, Cubasch H, Nietz S, Buccimazza I, Singh U, Cacala S, Stopforth L, Chen WC, McCormack VA, O'Neil DS, Jacobson JS, Neugut AI, Ruff P, Micklesfield LK (2019) The multimorbidity profile of South African women newly diagnosed with breast cancer. Int J Cancer. doi:10.1002/ijc.32727
101. Piette JD, Kerr EA (2006) The Impact of Comorbid Chronic Conditions on Diabetes Care. Diabetes Care 29 (3):725-731. doi:10.2337/diacare.29.03.06.dc05-2078
102. Niankara I, Niankara A (2017) The contribution of body mass index in the shared etiology of diabetes, hypertension and hyperlipidaemia: A semi-parametric trivariate probit modeling approach. Journal of Clinical Medicine and Therapeutics 2 (1):1-12
103. Ware LJ, Charlton K, Schutte AE, Cockeran M, Naidoo N, Kowal P (2017) Associations between dietary salt, potassium and blood pressure in South African adults: WHO SAGE

Wave 2 Salt & Tobacco. Nutr Metab Cardiovasc Dis 27 (9):784-791. doi:10.1016/j.numecd.2017.06.017
104. Ware LJ, Chidumwa G, Charlton K, Schutte AE, Kowal P (2019) Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2). Journal of human hypertension 33 (2):157-166
105. Vyas S, Kumaranayake L (2006) Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan 21 (6):459-468. doi:10.1093/heapol/czl029
106. Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography 38 (1):115-132
107. McKenzie D (2003) Measure inequality with asset indicators: Bureau for Research and Economic Analysis of Development. Centre for International Development, BREAD
108. Olaya B, Moneta MV, Caballero FF, Tyrovolas S, Bayes I, Ayuso-Mateos JL, Haro JM (2017) Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. BMC geriatrics 17 (1):186
109. Chang AY, Gómez-Olivé FX, Manne-Goehler J, Wade AN, Tollman S, Gaziano TA, Salomon JA (2019) Multimorbidity and care for hypertension, diabetes and HIV among older adults in rural South Africa. Bulletin of the World Health Organization 97 (1):10
110. Arokiasamy P, Uttamacharya U, Jain K, Biritwum RB, Yawson AE, Wu F, Guo Y, Maximova T, Espinoza BM, Salinas Rodríguez A, Afshar S, Pati S, Ice G, Banerjee S, Liebert MA, Snodgrass JJ, Naidoo N, Chatterji S, Kowal P (2015) The impact of multimorbidity on adult physical and mental health in low- and middle-income countries: what does the study on global ageing and adult health (SAGE) reveal? BMC Medicine 13 (1):178. doi:10.1186/s12916-015-0402-8
111. Islam MM, Valderas JM, Yen L, Dawda P, Jowsey T, McRae IS (2014) Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. PloS one 9 (1)
112. Whitson HE, Johnson KS, Sloane R, Cigolle CT, Pieper CF, Landerman L, Hastings SN (2016) Identifying patterns of multimorbidity in older Americans: application of latent class analysis. Journal of the American Geriatrics Society 64 (8):1668-1673
113. Mahomed OH, Asmall S, Freeman M (2014) An integrated chronic disease management model: a diagonal approach to health system strengthening in South Africa. Journal of health care for the poor and underserved 25 (4):1723-1729
114. Anand A, Sk MIK (2017) The risk of hypertension and other chronic diseases: Comparing smokeless tobacco with smoking. Frontiers in public health 5:255
115. Boffetta P, Straif K (2009) Use of smokeless tobacco and risk of myocardial infarction

and stroke: systematic review with meta-analysis. Bmj 339:b3060
116. Thankappan K, Thresia C (2007) Tobacco use & social status in Kerala. Indian Journal of Medical Research 126 (4):300
117. Fonda JR, Gregor KL, Fortier CB, Scioli-Salter ER, McGlinchey RE, Rasmusson A (2019) Tobacco dependence is associated with increased risk for multi-morbid clustering of posttraumatic stress disorder, depressive disorder, and pain among post-9/11 deployed veterans. Psychopharmacology 236 (6):1729-1739. doi:10.1007/s00213-018-5155-6
118. Gouda HN, Charlson F, Sorsdahl K, Ahmadzada S, Ferrari AJ, Erskine H, Leung J, Santamauro D, Lund C, Aminde LN (2019) Burden of non-communicable diseases in sub-Saharan Africa, 1990–2017: results from the Global Burden of Disease Study 2017. The Lancet Global Health 7 (10):e1375-e1387
119. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. Plos Med 3. doi:10.1371/journal.pmed.0030442
120. Pillay-van Wyk V, Msemburi W, Laubscher R, Dorrington RE, Groenewald P, Glass T, Nojilana B, Joubert JD, Matzopoulos R, Prinsloo M, Nannan N, Gwebushe N, Vos T, Somdyala N, Sithole N, Neethling I, Nicol E, Rossouw A, Bradshaw D (2016) Mortality trends and differentials in South Africa from 1997 to 2012: second National Burden of Disease Study. The Lancet Global Health 4 (9):e642-e653. doi:10.1016/S2214-109X(16)30113-9
121. Oni T, Youngblood E, Boulle A, McGrath N, Wilkinson RJ, Levitt NS (2015) Patterns of HIV, TB, and non-communicable disease multi-morbidity in peri-urban South Africa-a cross sectional study. BMC infectious diseases 15 (1):20
122. Bretzel RG (2007) Comorbidity of diabetes mellitus and hypertension in the clinical setting: A review of prevalence, pathophysiology, and treatment perspectives. Clinical Therapeutics 29:S35-S43. doi:https://doi.org/10.1016/j.clinthera.2007.07.010
123. Tarekegne FE, Padyab M, Schröders J, Williams JS (2018) Sociodemographic and behavioral characteristics associated with self-reported diagnosed diabetes mellitus in adults aged 50+ years in Ghana and South Africa: results from the WHO-SAGE wave 1. BMJ Open Diabetes Research and Care 6 (1)
124. Naude CE, Zani B, Ongolo-Zogo P, Wiysonge CS, Dudley L, Kredo T, Garner P, Young T (2015) Research evidence and policy: qualitative study in selected provinces in South Africa and Cameroon. Implementation Science 10 (1):126. doi:10.1186/s13012-015-0315-0
125. National Health Act 61 of 2003 (2003). South African National Department of Health
126. Lunn D, Spiegelhalter D, Thomas A, Best N (2009) The BUGS project: Evolution, critique and future directions. Stat Med 28 (25):3049-3067. doi:10.1002/sim.3680

127. Sesso HD, Cook NR, Buring JE, Manson JE, Gaziano JM (2008) Alcohol consumption and the risk of hypertension in women and men. Hypertension 51 (4):1080-1087. doi:10.1161/hypertensionaha.107.104968
128. Gatimu SM, Milimo BW, Sebastian MS (2016) Prevalence and determinants of diabetes among older adults in Ghana. BMC Public Health 16 (1):1174. doi:10.1186/s12889-016-3845-8
129. Ayah R, Joshi MD, Wanjiru R, Njau EK, Otieno CF, Njeru EK, Mutai KK (2013) A population-based survey of prevalence of diabetes and correlates in an urban slum community in Nairobi, Kenya. BMC Public Health 13 (1):1-11
130. Amoah AG, Owusu SK, Adjei S (2002) Diabetes in Ghana: a community based prevalence study in Greater Accra. Diabetes Research and Clinical Practice 56 (3):197-205
131. Wolf-Maier K, Cooper RS, Banegas JR, Giampaoli S, Hense H-W, Joffres M, Kastarinen M, Poulter N, Primatesta P, Rodríguez-Artalejo F (2003) Hypertension prevalence and blood pressure levels in 6 European countries, Canada, and the United States. The Journal of the American Medical Association 289 (18):2363-2369
132. Abera Abaerei A, Ncayiyana J, Levin J (2017) Health-care utilization and associated factors in Gauteng province, South Africa. Global Health Action 10 (1):1305765
133. Otwombe K, Dietrich J, Laher F, Hornschuh S, Nkala B, Chimoyi L, Kaida A, Gray GE, Miller CL (2015) Health-seeking behaviours by gender among adolescents in Soweto, South Africa. Global health action 8 (1):25670
134. Gerber LM, Stern PM (1999) Relationship of body size and body mass to blood pressure: sex-specific and developmental influences. Human Biology:505-528
135. Werfalli M, Kassanjee R, Kalula S, Kowal P, Phaswana-Mafuya N, Levitt NS (2018) Diabetes in South African older adults: prevalence and impact on quality of life and functional disability – as assessed using SAGE Wave 1 data. Global Health Action 11 (1):1449924. doi:10.1080/16549716.2018.1449924
136. Chidumwa G, Maposa I, Corso B, Minicuci N, Kowal P, Micklesfield LK, Ware LJ (2021) Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2. BMJ open 11 (1):e041604
137. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet 380 (9859):2095-2128
138. Jamison DT, Summers LH, Alleyne G, Arrow KJ, Berkley S, Binagwaho A, Bustreo F,

Evans D, Feachem RG, Frenk J (2013) Global health 2035: a world converging within a generation. The Lancet 382 (9908):1898-1955
139. Churchyard G, Mametja L, Mvusi L, Ndjeka N, Pillay Y, Hesseling A, Reid A, Babatunde S (2014) Tuberculosis control in South Africa: successes, challenges and recommendations: tuberculosis control-Progress towards the Millennium Development Goals. South African Medical Journal 104 (3):244-248
140. Gerrits MM, Van Oppen P, Van Marwijk HW, Van Der Horst H, Penninx BW (2013) The impact of chronic somatic diseases on the course of depressive and anxiety disorders. Psychotherapy and psychosomatics 82 (1):64-66
141. Hegeman JM, van Fenema EM, Comijs HC, Kok RM, van der Mast RC, de Waal MW (2017) Effect of chronic somatic diseases on the course of late-life depression. International journal of geriatric psychiatry 32 (7):779-787
142. Hsu W-C, Hsu H-C (2013) The effects of comorbidities on the trajectory of depressive symptoms among older adults in Taiwan. Journal of psychosomatic research 75 (5):414-418
143. Singer M, Bulled N, Ostrach B, Mendenhall E (2017) Syndemics and the biosocial conception of health. The Lancet 389 (10072):941-950
144. Ramlagan S, Peltzer K, Phaswana-Mafuya N (2013) Social capital and health among older adults in South Africa. BMC Geriatr 13:100. doi:10.1186/1471-2318-13-100
145. Armstrong T, Bull F (2006) Development of the world health organization global physical activity questionnaire (GPAQ). Journal of Public Health 14 (2):66-70
146. Andrews MR, Ceasar J, Tamura K, Langerman SD, Mitchell VM, Collins BS, Baumer Y, Gutierrez Huerta CA, Dey AK, Playford MP (2021) Neighborhood environment perceptions associate with depression levels and cardiovascular risk among middle-aged and older adults: Data from the Washington, DC cardiovascular health and needs assessment. Aging & mental health 25 (11):2078-2089
147. Garin N, Koyanagi A, Chatterji S, Tyrovolas S, Olaya B, Leonardi M, Lara E, Koskinen S, Tobiasz-Adamczyk B, Ayuso-Mateos JL (2016) Global multimorbidity patterns: a cross-sectional, population-based, multi-country study. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences 71 (2):205-214
148. Chang AY, Gómez-Olivé FX, Payne C, Rohr JK, Manne-Goehler J, Wade AN, Wagner RG, Montana L, Tollman S, Salomon JA (2019) Chronic multimorbidity among older adults in rural South Africa. BMJ Global Health 4 (4):e001386
149. Gamma A, Angst J (2001) Concurrent psychiatric comorbidity and multimorbidity in a community study: gender differences and quality of life. European Archives of Psychiatry and Clinical Neuroscience 251 (2):43-46

150. Kruger HS (2018) Obesity among women: A complex setting. South African Journal of Clinical Nutrition 31 (4):4-5
151. Millar WJ, Wigle DT (1986) Socioeconomic disparities in risk factors for cardiovascular disease. CMAJ: Canadian Medical Association Journal 134 (2):127
152. Marmot M, Friel S (2008) Global health equity: evidence for action on the social determinants of health. Journal of Epidemiology & Community Health 62 (12):1095-1097
153. Michael YL, Berkman LF, Colditz GA, Holmes MD, Kawachi I (2002) Social networks and health-related quality of life in breast cancer survivors: a prospective study. Journal of Psychosomatic research 52 (5):285-293
154. Kroenke CH, Kubzansky LD, Schernhammer ES, Holmes MD, Kawachi I (2006) Social networks, social support, and survival after breast cancer diagnosis. Journal of clinical oncology 24 (7):1105-1111
155. Tinetti ME, Fried TR, Boyd CM (2012) Designing health care for the most common chronic condition—multimorbidity. Jama 307 (23):2493-2494
156. Pahl R (2003) Some sceptical comments on the relationship between social support and well-being. Leisure studies 22 (4):357-368
157. Ghose B, Abdoul Razak MY (2017) Memory and learning complaints in relation to depression among elderly people with multimorbidity. Geriatrics 2 (2):15
158. Voinov B, Richie WD, Bailey RK (2013) Depression and chronic diseases: it is time for a synergistic mental health and primary care approach. Prim Care Companion CNS Disord 15 (2):8
159. Jackson CA, Sudlow CL, Mishra GD (2018) Psychological distress and risk of myocardial infarction and stroke in the 45 and up study: A prospective cohort study. Circulation: Cardiovascular Quality and Outcomes 11 (9):e004500
160. Cohen BE, Edmondson D, Kronish IM (2015) State of the art review: depression, stress, anxiety, and cardiovascular disease. American journal of hypertension 28 (11):1295-1302
161. Wang JL, Schmitz N, Dewa CS (2010) Socioeconomic status and the risk of major depression: the Canadian National Population Health Survey. Journal of Epidemiology and Community Health 64 (5):447-452. doi:10.1136/jech.2009.090910
162. Moffat K, Mercer SW (2015) Challenges of managing people with multimorbidity in today's healthcare systems. BMC family practice 16 (1):1-3
163. Sahoo KC, Kanungo S, Mahapatra P, Pati S (2021) Non-communicable diseases care during COVID-19 pandemic: A mixed-method study in Khurda district of Odisha, India. Indian J Med Res 153 (5&6):649-657. doi:10.4103/0971-5916.323435
164. Organization WH (2020) The impact of the COVID-19 pandemic on noncommunicable

disease resources and services: results of a rapid assessment.
165. English PB (2001) An Introductory Guide to Disease Mapping. American Journal of Epidemiology 154 (9):881-882. doi:10.1093/aje/154.9.881-a
166. Puoane TR, Tsolekile LP, Caldbick S, Igumbor EU, Meghnath K, Sanders D (2012) Chronic non-communicable diseases in South Africa: Progress and challenges. South African Health Review 2012 (1):115-126
167. D'Haese M, Van Huylenbroeck G (2005) The rise of supermarkets and changing expenditure patterns of poor rural households case study in the Transkei area, South Africa. Food Policy 30 (1):97-113
168. Temple NJ, Steyn NP, Fourie J, De Villiers A (2011) Price and availability of healthy food: a study in rural South Africa. Nutrition 27 (1):55-58
169. Odunitan-Wayas F, Okop K, Dover R, Alaba O, Micklesfield L, Puoane T, Uys M, Tsolekile L, Levitt N, Battersby J (2018) Food purchasing characteristics and perceptions of neighborhood food environment of South Africans living in low-, middle-and high-socioeconomic neighborhoods. Sustainability 10 (12):4801
170. Odunitan-Wayas FA, Okop KJ, Dover RV, Alaba OA, Micklesfield LK, Puoane T, Levitt NS, Battersby J, Meltzer ST, Lambert EV (2021) Food purchasing behaviour of shoppers from different South African socio-economic communities: results from grocery receipts, intercept surveys and in-supermarkets audits. Public Health Nutrition 24 (4):665-676
171. Ware LJ, Charlton K, Schutte AE, Cockeran M, Naidoo N, Kowal P (2017) Associations between dietary salt, potassium and blood pressure in South African adults: WHO SAGE Wave 2 Salt & Tobacco. Nutrition, Metabolism and Cardiovascular Diseases 27 (9):784-791. doi:10.1016/j.numecd.2017.06.017
172. Patel V, Chatterji S (2015) Integrating mental health in care for noncommunicable diseases: an imperative for person-centered care. Health Affairs 34 (9):1498-1505

APPENDICES

Appendix A: Summary of NCD's missing data.

Variable	Missing (n)	Total (N)	Percent Missing
Arthritis	403	3180	12.7
Angina	403	3180	12.7
Diabetes	404	3180	12.7
Chronic lung disease	88	3180	2.8
Asthma	112	3180	3.5
Depression	405	3180	12.7
Hypertension	406	3180	12.8

Appendix B: Syntax for Paper 1.

SAS code

```
proc import out= Final_analytical datafile = "E:\PhD\Papers\Paper 1\P1_analytical.dta"
DBMS=STATA REPLACE;
run;
proc contents data=Final_analytical;
run;
* LCA with 2 classes, taking into account for age groups;
proc lca data = Final_analytical OUTPARAM = test outpost = P1_post;
        nclass 2;
        id id;
        items Arthritis Stroke Angina Diabetes Lung Asthma Depression HTN ;
        categories  2 2 2 2 2 2 2 2 ;
        seed 123741;
        group Age_cat;
run;
* LCA with 3 classes, taking into account for age groups;
proc lca data = Final_analytical OUTPARAM = test outpost = P2_post;
        nclass 3;
        id id;
        items Arthritis Stroke Angina Diabetes Lung Asthma Depression HTN ;
        categories  2 2 2 2 2 2 2 2 ;
        seed 123741;
        rho prior=1;
        group Age_cat;
run;
* LCA with 4 classes, taking into account for age groups;
proc lca data = Final_analytical OUTPARAM = test outpost = P3_post;
        nclass 4;
        id id;
        items Arthritis Stroke Angina Diabetes Lung Asthma Depression HTN ;
        categories 2 2 2 2 2 2 2 2 ;
        seed 123741;
```

```
        rho prior=1;
        group Age_cat;
run;
* LCA with 5 classes, taking into account for age groups;
proc lca data = Final_analytical OUTPARAM = test outpost = P4_post;
        nclass 5;
        id id;
        items Arthritis Stroke Angina Diabetes Lung Asthma Depression HTN ;
        categories 2 2 2 2 2 2 2 2 ;
        seed 123741;
        rho prior=1;
        group Age_cat;
run;
* LCA with 6 classes, taking into account for age groups;
proc lca data = Final_analytical OUTPARAM = test outpost = P5_post;
        nclass 6;
        id id;
        items Arthritis Stroke Angina Diabetes Lung Asthma Depression HTN ;
        categories 2 2 2 2 2 2 2 2 ;
        seed 123741;
        rho prior=1;
        group Age_cat;
run;
```

Stata code

```
use "E:\PhD\Datasets\Wave 2\ZAF_W2_with_LCA copy.dta", clear
cd "E:\PhD\Papers\Paper 1"
/*
rename q3007 Alcohol
rename q3001 Tobbaco
rename q3022 Active_transport
rename q0104 Residence
rename q1503 Currently_working
rename q1017 Years_educated
rename q1012 M_status
ren q1009 Sex
rename best Latent_classes
ren q1018 Ethnicity
gen adjpweight=pweight/24766.2 if age_cat==1
replace adjpweight=pweight/4662.422 if age_cat==2
replace q2506=. if q2506<120 | q2506>200  // Q2506: True height (cm) - invalid values (<120 or >200cm) removed
replace q2507=. if q2507<35 | q2507>180   // Q2507: True weight (kg) - invalid values (<35 or >180kg) removed

replace q2508=. if q2508<40 | q2508>170  // Q2508: Waist Circumference centimeters - invalid values (<40 or >170cm) removed
replace q2509=. if q2509<50 | q2509>190  // Q2509: HC centimeters - invalid values (<50 or >190cm) removed
ren q2507 Weight_kg
ren q2506 Height_cm
ren q2508 WC
ren q2509 HC
gen Height_m= Height_cm/100
gen BMI=Weight_kg/Height_m^2
sum BMI
replace BMI=. if BMI<13 | BMI>70  // Weight (kg) / height (m)2 - invalid values (<13 or >70) removed
gen WHtR= WC/Height_cm
```

**********Table 1

rename (multimorbidity hypertension depression diabetes lung asthma vision arthritis stroke angina) (Multimorbidity Hypertension Depression Diabetes Lung Asthma Vision Arthritis Stroke Angina)

**ssc install labutil2 ///How to assign variable names as their labels

labvars Multimorbidity Hypertension Depression Diabetes Lung Asthma Vision Arthritis Stroke Angina, names

Else: How to rename variables according to their labels

ssc install labutil2

lab2varn [varlist]

replace Multimorbidity=2 if Multimorbidity==0

label define yesno 1 "Yes" 2 "No"

foreach v of varlist Multimorbidity Hypertension Depression Diabetes Lung Asthma Vision Arthritis Stroke Angina {

label values `v' yesno

}

recode latent_classes (2=0 "Healthy") (1=1 "Hypertensive/Diabetic") (3=2 "Other NCDs"), gen(Classes)

label define age 1 "18 to 49" 2 "50 +"

lab val age_cat age

gen Exercise=1 if q3016==1 | q3025==1

replace Exercise=0 if Exercise!=1

replace Exercise=. if q3016==. | q3025==.

recode q7521 (1 2=1 "Good/Very good") (3=2 "Moderate") (4 5=3 "Poor/Very poor") (9=.), gen(QoSleep)

recode QoSleep (1 = 1 "Good") (2 3 =2 "Bad"), gen(Sleep_quality)

ren q3015a salt_at_table

recode salt_at_table (1/4=1 "Yes") (5=2 "No"), gen(addsaltattable)

labvars sex Ethnicity alcohol tobbaco addsaltattable Exercise residence currently_working Sleep_quality Age BMI WC HC WHtR Years_educated, names

*/

summtab, catvars(Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina) by(age_cat) word wordname(Table1_by_age_weighted) catptype(1) total

```
title(Table 1: Summary statistics for chronic conditions, by age groups) wts(adjpweight)
replace
*### Table 1**
summtab, cat_vars(Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis
Stroke Angina) by(age_cat) excel excelname(Table1_by_age_weighted) cat_ptype(1) total
title(Table 1: Summary statistics for chronic conditions) wts(adjpweight) replace
*### Table 1**
summtab, cat_vars(sex Ethnicity alcohol tobbaco addsaltattable Exercise residence
currently_working Sleep_quality) cont_vars(Age BMI WC HC WHtR Years_educated)
by(Classes) word wordname(Table1_by_latentclasses) median mnfmt(2)
cont_ptype(1)cat_ptype(1) pval title(Table 1: Characteristics of participants by latent class
category (n = 1967)) replace
bysort tab1 Multimorbidity Hypertension Depression Diabetes Lung Asthma Vision Arthritis
Angina [aw=adjpweight]

bysort latent_classes:tab1 sex  [aw=adjpweight]
//Ologit
svyset q0101b [pw=adjpweight] , strata(strata)
ta strata Classes
svy: ologit Classes i.sex i.Ethnicity ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise
i.residence ib2.currently_working i.Sleep_quality Age BMI WHtR Years_educated if
age_cat==1 & strata!=5 & strata!=12 & strata!=17 & strata!=18, nolog or
svy: ologit Classes i.sex i.Ethnicity ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise
i.residence ib2.currently_working i.Sleep_quality Age BMI WHtR Years_educated if
age_cat==2 & strata!=5 & strata!=12 & strata!=17 & strata!=18, nolog or
svy: mlogit Classes i.sex i.Ethnicity ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise
i.residence ib2.currently_working i.Sleep_quality Age BMI Years_educated if strata!=5 &
strata!=12 & strata!=17 & strata!=18, base(0) rrr
//brant, detail
***Assumption doesnt hold:p=.02***
//Mlogit
mlogit Classes Age i.sex i.Ethnicity i.alcohol i.tobbaco i.addsaltattable i.Exercise i.residence
i.currently_working i.Sleep_quality  BMI WHtR Years_educated, base(0) rrr
outreg2 using "E:\PhD\Papers\Paper 1\Multinomial.doc", stats(coef ci) append eform
*###For regression###*
```

```
 ///Replacing label "2" as "0" for NO
 foreach v of varlist Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina
{
replace `v'=0 if `v'==2
 }
  //////7 condtions/////////
 clear
 use "E:\PhD\Datasets\Wave 2\Analytical_no_stroke.dta", clear
 rename q3007 Alcohol
rename q3001 Tobbaco
rename q3022 Active_transport
rename q0104 Residence
rename q1503 Currently_working
rename q1017 Years_educated
rename q1012 M_status
ren q1009 Sex
rename best Latent_classes
ren q1018 Ethnicity
//gen adjpweight=pweight/24766.2 if age_cat==1
replace adjpweight=pweight/4662.422 if age_cat==2
replace q2506=. if q2506<120 | q2506>200  // Q2506: True height (cm) - invalid values (<120 or >200cm) removed
replace q2507=. if q2507<35 | q2507>180   // Q2507: True weight (kg) - invalid values (<35 or >180kg) removed
replace q2508=. if q2508<40 | q2508>170  // Q2508: Waist Circumference centimeters - invalid values (<40 or >170cm) removed
replace q2509=. if q2509<50 | q2509>190  // Q2509: HC centimeters - invalid values (<50 or >190cm) removed
ren q1011 Age
ren age_cat Age_act
ren q2507 Weight_kg
ren q2506 Height_cm
ren q2508 WC
ren q2509 HC
ren Age_act Age_cat
```

```
gen Height_m= Height_cm/100
gen BMI=Weight_kg/Height_m^2
sum BMI
replace BMI=. if BMI<13 | BMI>70  // Weight (kg) / height (m)2 - invalid values (<13 or >70) removed
gen WHtR= WC/Height_cm
**********Table 1
ren hypertension htn_sr
ren hyper_bp_medic hypertension
ren multimorbidity mm_sr
ren multimorbidity2 multimorbidity

rename (multimorbidity hypertension depression diabetes lung asthma arthritis stroke angina) (Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina)
**ssc install labutil2 ///How to assign variable names as their labels
labvars Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina Hypertension, names
//Else: How to rename variables according to their labels
//ssc install labutil2
//lab2varn [varlist]
replace Multimorbidity=2 if Multimorbidity==0
label define yesno 1 "Yes" 2 "No"
foreach v of varlist Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina {
label values `v' yesno
 }
 recode Latent_classes (3=0 "Healthy") (2=1 "Early risk") (1=2 "High MM"), gen(Classes)
label define Agecat 1 "18 to 49" 2 "50 +"
lab val Age_cat Agecat
gen Exercise=1 if q3016==1 | q3025==1
replace Exercise=0 if Exercise!=1
replace Exercise=. if q3016==. | q3025==.
recode q7521 (1 2=1 "Good/Very good") (3=2 "Moderate") (4 5=3 "Poor/Very poor") (9=.), gen(QoSleep)
recode QoSleep (1 = 1 "Good") (2 3 =2 "Bad"), gen(Sleep_quality)
```

```
ren q3015a salt_at_table
recode salt_at_table (1/4=1 "Yes") (5=2 "No"), gen(Addsaltattable)
labvars Sex Ethnicity Alcohol Tobbaco Addsaltattable Exercise Residence Currently_working Sleep_quality Age BMI WC HC WHtR Years_educated, names
 cd "E:\PhD\Papers\Paper 1"
bysort Age_cat: tab1 Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina [aw=adjpweight]
summtab, cat_vars(Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina) by(Age_cat) excel excelname(Table1_by_age_weighted) cat_ptype(1) total title(Table 1: Summary statistics for chronic conditions) wts(adjpweight) replace
summtab if Age_cat==1, cat_vars(Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina) by(Sex) excel excelname(Table1_by_sex_18to49_weighted) cat_ptype(1) total title(Table 1: Summary statistics for chronic conditions) wts(adjpweight) replace
summtab if Age_cat==2, cat_vars(Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis Stroke Angina) by(Sex) excel excelname(Table1_by_sex_50+_weighted) cat_ptype(1) total title(Table 1: Summary statistics for chronic conditions) wts(adjpweight) replace
tab Classes [aw=adjpweight]
summtab, cat_vars(Sex Ethnicity Alcohol Tobbaco Addsaltattable Exercise Residence Currently_working Sleep_quality) cont_vars(Age BMI WC HC WHtR Years_educated) by(Classes) word wordname(Table3_by_latentclasses) median mnfmt(2) cont_ptype(1)cat_ptype(1) total title(Table 1: Characteristics of participants by latent class category (n = 2761)) replace
svyset q0101b [pw= adjpweight], strata(strata)
svy: mlogit Classes Age i.Sex i.Ethnicity ib2.Alcohol ib2.Tobbaco ib2.Addsaltattable i.Exercise i.Residence ib2.Currently_working i.Sleep_quality BMI WHtR Years_educated if Age_cat==1 & strata!=5 & strata!=12 & strata!=17 & strata!=18, base(0) rrr
//svy: mlogit Classes Age i.Sex i.Ethnicity ib2.Alcohol ib2.Tobbaco ib2.Addsaltattable i.Exercise i.Residence ib2.Currently_working i.Sleep_quality BMI WHtR Years_educated if Age_cat==1 & strata!=5 & strata!=12 & strata!=17 & strata!=18& Sex==1, base(0) rrr
//svy: mlogit Classes Age i.Sex i.Ethnicity ib2.Alcohol ib2.Tobbaco ib2.Addsaltattable i.Exercise i.Residence ib2.Currently_working i.Sleep_quality BMI WHtR Years_educated if Age_cat==1 & strata!=5 & strata!=12 & strata!=17 & strata!=18& Sex==2, base(0) rrr
svy: mlogit Classes Age i.Sex i.Ethnicity ib2.Alcohol ib2.Tobbaco ib2.Addsaltattable
```

```
i.Exercise i.Residence ib2.Currently_working i.Sleep_quality BMI WHtR Years_educated if Age_cat==2 & strata!=5 & strata!=12 & strata!=17 & strata!=18, base(0) rrr
//svy: mlogit Classes Age i.Sex i.Ethnicity ib2.Alcohol ib2.Tobbaco ib2.Addsaltattable i.Exercise i.Residence ib2.Currently_working i.Sleep_quality BMI WHtR Years_educated if Age_cat==2 & strata!=5 & strata!=12 & strata!=17 & strata!=18 & Sex==1, base(0) rrr
//svy: mlogit Classes Age i.Sex i.Ethnicity ib2.Alcohol ib2.Tobbaco ib2.Addsaltattable i.Exercise i.Residence ib2.Currently_working i.Sleep_quality BMI WHtR Years_educated if Age_cat==2 & strata!=5 & strata!=12 & strata!=17 & strata!=18 & Sex==2, base(0) rrr
/*
ren classes Classes
ren age Age
ren bmi BMI
ren wc WC
ren hc HC
ren years_educated Years_educated
ren ethnicity Ethnicity
ren sleep_quality Sleep_quality
ren whtr WHtR
ren exercise Exercise
ren (multimorbidity hypertension depression diabetes lung asthma vision arthritis stroke angina) (Multimorbidity Hypertension Depression Diabetes Lung Asthma Vision Arthritis Stroke Angina)
labvars Multimorbidity Hypertension Depression Diabetes Lung Asthma Vision Arthritis Stroke Angina, names
*/
///7 April 2019
use "E:\PhD\Datasets\Wave 2\SouthAfricaHHDataW2.dta", clear
keep hhid q0002 q0508 q0401 q0510 q0701- q0719
rename q0002 q0002_original
gen q0002=substr(hhid,4,5)
sort q0002
save "E:\PhD\Papers\Paper 1\Assets.dta", replace
use "E:\PhD\Papers\Paper 1\Analytical_no_stroke", clear
sort q0002
tostring q0002, replace
```

```
sort q0002
merge q0002 using "E:\PhD\Papers\Paper 1\Assets.dta", update replace
ta _merge
keep if _merge==3
drop _merge
saveold "E:\PhD\Papers\Paper 1\Analytical_no_stroke", version(13) replace
*/

pca q0701 q0702 q0703 q0704 q0705 q0706 q0707 q0708 q0709 q0710 q0711 q0712 q0713
q0714 q0715 q0716 q0717 q0718 q0719 q0508 q0401 q0510
predict wealthindex
codebook wealthindex
*xtile quintiles = wealthindex , nq(5)
*ta quintiles
xtile tertiles = wealthindex , nq(3)
ta tertiles
bysort sex: tab1 Multimorbidity Hypertension Depression Diabetes Lung Asthma Arthritis
Angina [aw=adjpweight]
summtab, cat_vars(sex Ethnicity Alcohol Tobbaco Addsaltattable Exercise Residence
Sleep_quality) cont_vars(Age BMI WC HC WHtR Years_educated) by(Classes) word
wordname(Table3_by_latentclasses) median mnfmt(2) cont_ptype(1)cat_ptype(1) total
title(Table 1: Characteristics of participants by latent class category (n = 1967)) replace
svyset q0101b [pw= adjpweight], strata(strata)
*svy: mlogit Classes i.sex ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise i.residence
i.quintiles i.Sleep_quality Age BMI Years_educated if strata!=5 & strata!=12 & strata!=17 &
strata!=18, base(0) rrr
svy: mlogit Classes i.sex ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise i.residence
i.tertiles i.Sleep_quality Age BMI Years_educated if strata!=5 & strata!=12 & strata!=17 &
strata!=18, base(0) rrr
tab Classes [aw=adjpweight]
*tab quintiles Classes [aw=adjpweight], row
tab tertiles Classes [aw=adjpweight], row
tab sex [aw=adjpweight]
tab Ethnicity [aw=adjpweight]
dunntest age, by(Classes) ma(bonferroni) nokwallis
```

```
dunntest BMI , by(Classes) ma(bonferroni) nokwallis
dunntest WC , by(Classes) ma(bonferroni) nokwallis
dunntest HC , by(Classes) ma(bonferroni) nokwallis
dunntest WHtR , by(Classes) ma(bonferroni) nokwallis
dunntest Years_educated , by(Classes) ma(bonferroni) nokwallis
tabi 545 51 \ 1046 197, chi2
tabi 1046 197 \ 27 101, chi2
tabi 51 27\ 197 101, chi2
tabi 545 27 \ 1046 101, chi2
tabi 289 31 \ 1296 214 , chi2
tabi 31 29 \ 214 98 , chi2
tabi 289 29 \ 1296 98 , chi2
tabi 301 35 \1284 210 , chi2
tabi 301 40 \ 1284 86, chi2
tabi 35 40 \ 210 86, chi2
tabi 1084 155 \ 501 90, chi2
tabi 155 73 \90 55 , chi2
tabi 1084 73 \ 501 55, chi2
tabi 1396 218 \ 181 26, chi2
tabi 218 108 \26 20, chi2
tabi 1396 108 \ 181 20, chi2
tabi 1124 160 \ 467 88, chi2
tabi 1124 101 \ 467 27, chi2
tabi 160 101 \ 88 27, chi2
tabi 1307 176\263 65, chi2
tabi 1307 89\263 36, chi2
tabi 176 89\65 36, chi2
*******
ren Exercise exercise1
gen Exercise=1 if q3016==1 | q3025==1 | q3019==1 | q3028==1
replace Exercise=0 if Exercise!=1
replace Exercise=. if q3028==. | q3025==. | q3016==. | q3019==.
svy: mlogit Classes i.sex ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise i.residence
i.tertiles i.qosleep Age BMI Years_educated if strata!=5 & strata!=12 & strata!=17 &
strata!=18, base(0) rrr
```

/////13 October 2020

recode BMI (0/18.49999999=1 "Underweight") (18.5/24.9999999=2 "Healthy weight") (25/29.999=3 "Overweight") (30/1000=4 "Obese"), gen(BMI_cat)

bys Classes: tab1 Hypertension Depression Diabetes Lung Asthma Arthritis Angina

Appendix C: Syntax for Paper 2

OpenBugs code

```
##############################################
########## Shared Component model ###############
##############################################
model
{
#likelihood
for(i in 1: N)
{
#for Hypertension
Hypertension[i]~dbern(p1[i])
p1[i]<-min(1,max(0,pHypertension[i]))
logit(pHypertension[i])<-alpha+beta1*Age[i]+
beta2[sex[i]]+beta3[tertiles[i]]+beta4*Years_educated[i]+beta5*BMI[i]+beta6[grant[i]]+beta
7[addsaltattable[i]]+S[1,district[i]]
#for Diabetes
Diabetes[i]~dbern(p2[i])
p2[i]<-min(1,max(0,pDiabetes[i]))
logit(pDiabetes[i])<-neta+beta8*Age[i]+
beta9[sex[i]]+beta10[tertiles[i]]+beta11*Years_educated[i]+beta12[tobbaco[i]]+S[2,district[i]
]
}
alpha~dnorm(0.0,1.0E-4) #prior for intercept
beta1~dnorm(0.0,1.0E-4) #prior for beta1
beta2[1]<-0 #reference category
beta2[2]~dnorm(0.0, 1.0E-4) #prior for beta2 cat 2
beta3[1]<-0 #reference category
beta3[2]~dnorm(0.0, 1.0E-4) #prior for beta3 cat 2
beta3[3]~dnorm(0.0, 1.0E-4) #prior for beta3 cat 3
beta4~dnorm(0.0, 1.0E-4) #prior for beta4
beta5~dnorm(0.0, 1.0E-4) #prior for beta5
beta6[1]<-0 #reference category
beta6[2]~dnorm(0.0, 1.0E-4) #prior for beta6 cat 2
beta7[1]<-0 #reference category
```

```
beta7[2]~dnorm(0.0, 1.0E-4) #prior for beta7 cat 2
neta~dnorm(0.0,1.0E-4) #prior for intercept
beta8~dnorm(0.0,1.0E-4) #prior for beta8
beta9[1]<-0 #reference category
beta9[2]~dnorm(0.0, 1.0E-4) #prior for beta9 cat 2
beta10[1]<-0 #reference category
beta10[2]~dnorm(0.0, 1.0E-4) #prior for beta10 cat 2
beta10[3]~dnorm(0.0, 1.0E-4) #prior for beta10 cat 3
beta11~dnorm(0.0, 1.0E-4) #prior for beta11
beta12[1]<-0 #reference category
beta12[2]~dnorm(0.0, 1.0E-4) #prior for beta12 cat 2
#Getting Odds ratio from logOdds, by taking exponent of the coefficients
#for(i in 3:10)
#{
#Oddsbeta1[i]<-exp(beta1[i]); Oddsbeta2[i]<-exp(beta2[i]); Oddsbeta3[i]<-exp(beta3[i]);
#Oddsbeta4[i]<-exp(beta4); Oddsbeta5[i]<-exp(beta5[i]); Oddsbeta6[i]<-exp(beta6[i]);
#Oddsbeta7[i]<-exp(beta7[i]); Oddsbeta8[i]<-exp(beta8)
#}
for(i in 1:Nareas)
{
S[1, i] <- phi[i] * delta + psi[1, i]
S[2, i] <- phi[i] / delta + psi[2, i]
}
# Spatial priors (BYM) for the disease-specific random effects
for (k in 1 : Ndiseases) {
for (i in 1 : Nareas) {
# convolution prior = sum of unstructured and spatial effects
psi[k, i] <- U.sp[k, i] + S.sp[k, i]
# unstructured disease-specific random effects
U.sp[k, i] ~ dnorm(0, tau.unstr[k])
}
# spatial disease-specific effects
S.sp[k,1:Nareas] ~ car.normal(adj[],weights[], num[],tau.spatial[k])
}
# Spatial priors (BYM) for the shared random effects
```

```
for (i in 1:Nareas) {
# convolution prior = sum of unstructured and spatial effects
phi[i] <- U.sh[i] + S.sh[i]
# unstructured shared random effects
U.sh[i] ~ dnorm(0, omega.unstr)
}
# spatial shared random effects
S.sh[1:Nareas] ~ car.normal(adj[], weights[], num[], omega.spatial)
for (k in 1:sumNumNeigh) [31]
#prior
for(i in 1: N)
{
for(j in 1: 52)
{
PHTN[j,i]<-(pHypertension[i])*(equals(district[i],j))
PDBTS[j,i]<-(pDiabetes[i])*(equals(district[i],j))
}
}
for(j in 1: 52)
{
for(i in 1: N)
{
count[j,i]<-equals(district[i],j)
}
number[j]<-sum(count[j,])
PH[j]<-sum(PHTN[j,])/number[j]
PD[j]<-sum(PDBTS[j,])/number[j]
}
# Other priors
for (k in 1:Ndiseases) {
tau.unstr[k] ~ dgamma(0.5, 0.0005)
tau.spatial[k] ~ dgamma(0.5, 0.0005)
}
omega.unstr ~ dgamma(0.5, 0.0005)
omega.spatial ~ dgamma(0.5, 0.0005)
```

```
logdelta ~ dnorm(0, 5.9)
delta<-exp(logdelta)
###OR.ratio<-pow(delta, 2)
        # Relative risks and other summary quantities
                # The GeoBUGS map tool can only map vectors, so need to create
separate vector
                # of quantities to be mapped, rather than an array (i.e. totalRR[i,k]
won't work!)
                        for (i in 1 : Nareas) {
                                                totalRR1[i] <- exp(S[1,i])
        # overall RR of disease 1 (hypertension) in area i
                                        totalRR2[i] <- exp(S[2,i])                      #
overall RR of disease 2 (diabetes) in area i
                # residulal RR specific to disease 1 (oral cancer)
                                        specificRR1[i]<- exp(psi[1,i])
                # residulal RR specific to disease 2 (lung cancer)
                                        specificRR2[i]<- exp(psi[2,i])
                        # shared component of risk common to both diseases
                                        sharedRR[i] <- exp(phi[i])
                # Note that this needs to be scaled by delta or 1/delta if the
                # absolute magnitude of shared RR for each disease is of interest
                                        logsharedRR1[i] <- phi[i] * delta
                                        logsharedRR2[i] <- phi[i]  /delta
                                }
                # empirical variance of shared effects (scaled for disease 1)
                        var.shared[1] <- sd(logsharedRR1[])*sd(logsharedRR1[])
                # empirical variance of shared effects  (scaled for disease 2)
                                var.shared[2] <-
sd(logsharedRR2[])*sd(logsharedRR2[])
                # empirical variance of disease 1 specific effects
                                var.specific[1] <- sd(psi[1,])*sd(psi[1,])
                # empirical variance of disease 2 specific effects
                                var.specific[2] <- sd(psi[2,])*sd(psi[2,])
                                # fraction of total variation in relative risks for each
disease that is explained
```

```
                    # by the shared component
                        #frac.shared[1] <- var.shared[1] / (var.shared[1] +
var.specific[1])
                        #frac.shared[2] <- var.shared[2] / (var.shared[2] +
var.specific[2])

}
#DATA
#INITIALS
list(alpha=0, beta1=0.4, beta2=c(NA, 0.2), beta3=c(NA, 0.001, 0.28), beta4=0.8, beta5=0.8,
beta6=c(NA, 0.1), beta7=c(NA, 0.39), neta=0, beta8=0.25, beta9=c(NA, 1.7), beta10=c(NA,
0.7, 0.3), beta11=0.7, beta12=c(NA, 1.8), omega.unstr=9.4,omega.spatial=4.0,logdelta=2,
tau.unstr = c(0.25, 2.7), tau.spatial = c(7.5, 9.8))
list(alpha=1, beta1=0.3, beta2=c(NA, 0.22), beta3=c(NA, 0.01, 0.72), beta4=0.97, beta5=0.1,
beta6=c(NA, 0.55), beta7=c(NA, 0.3), neta=0, beta8= 0.35, beta9=c(NA, 0.07), beta10=c(NA,
0.8, 0.56), beta11=0.2, beta12=c(NA, 2.8),
omega.unstr=0.41,omega.spatial=0.001,logdelta=0.5,
tau.unstr=c(3.5,2),tau.spatial=c(3.0,1.5))
list(alpha=2, beta1=0.37, beta2=c(NA, 0.8), beta3=c(NA, 0.1, 0.93), beta4=0.01, beta5=0.008,
beta6=c(NA, 0.27), beta7=c(NA, 0.09), neta=0, beta8=0.7, beta9=c(NA, 1.27), beta10=c(NA,
0.2, 0.73), beta11=0.86, beta12=c(NA, 0.262),
omega.unstr=1.0,omega.spatial=1.0,logdelta=1.4,
tau.unstr=c(2.0,1.0),tau.spatial=c(1.0,1.0))
```

Stata code

```
use "E:\PhD\Papers\Paper 2\Data.dta"
destring id, replace force
ren prov Province
merge 1:1 id using "E:\PhD\Papers\Paper 2\Prov codes.dta"
keep id residence sex age yearseducated currentlyworking tobbaco alcohol arthritis diabetes hypertension multimorbidity classes exercise qosleep whtr addsaltattable prov prov1 tertiles
////Keep only complete cases
**findit rmiss2

egen nmis = rmiss2(id residence sex age yearseducated currentlyworking tobbaco alcohol arthritis diabetes hypertension multimorbidity classes exercise qosleep whtr addsaltattable prov tertiles)
keep if (nmis == 0)
/////
save "E:\PhD\Papers\Paper 2\Complete data.dta", replace
///Save as CSV with no labels
export delimited using "E:\PhD\Papers\Paper 2\Complete data.csv", nolabel replace
/////Describing misssingness
mdesc
///// 2 September 2019
use "E:\PhD\Papers\Paper 2\Analytical data _ 2 Sept 20.dta"
drop  _merge
merge m:1 HHid using "E:\PhD\Papers\Paper 2\Assets and grant.dta"
keep if _merge==3
cd "E:\PhD\Papers\Paper 2"
ren q0607 grant
replace grant=. if grant==8
recode q1016 (4 5 6=1 "Yes") (1 2 3=0 "No"), gen(Completed_HS)
gen medical_aid=.
tab1 q5010_6 q5010_7 // Only 4 with voluntary mediacal aid
recode Age (0/34.9999999=1 "18 - 34") (35/1000=2 "35 and above"), gen(Age_cat)
gen Comorbidity=.
replace Comorbidity=1 if Hypertension==1 & Diabetes==1 & Hypertension!=. & Diabetes!=.
replace Comorbidity=0 if Comorbidity!=1
```

```
gen BP=(q2501_s+q2502_s+q2503_s)/3
gen DBP=( q2501_d + q2502_d + q2503_d )/3
ren BP SBP
gen Normotensive=.
replace Normotensive=1 if SBP<120 | DBP<80
replace Normotensive=0 if Normotensive!=1
gen Pre_hypertensive=.
replace Pre_hypertensive=1 if SBP>=120 & SBP<=139.99 | DBP>=80 & DBP<=89.99
replace Pre_hypertensive=0 if Pre_hypertensive!=1
gen Hypertensive=.
replace Hypertensive=1 if SBP>=140 | DBP >=90
replace Hypertensive=0 if Hypertensive!=1
recode BMI (0/18.49999999=1 "Underweight") (18.5/24.9999999=2 "Healthy weight")
(25/29.999=3 "Overweight") (30/1000=4 "Obese"), gen(BMI_cat)
swilk WHtR Age Years_educated
summtab, by(Comorbidity) catvars(Age_cat sex currently_working Exercise residence
Completed_HS tertiles grant BMI_cat Normotensive Pre_hypertensive Hypertensive alcohol
tobbaco) contvars(WHtR Years_educated) word wordname(Table1_by_Comorbidity)
catptype(1) contptype(2) median total pval title("Table 1: Summary statistics for
demographic, socioeconomic status, anthropometry and blood pressure characteristics, by
Comorbidity") replace
stepwise, pr(.2): logit Comorbidity sex currently_working Exercise residence Completed_HS
tertiles grant Age_cat BMI_cat Years_educated alcohol tobbaco
stepwise, pr(.1): logit Hypertension sex Exercise residence  tertiles grant Age BMI
Years_educated alcohol tobbaco addsaltattable
logistic Hypertension i.sex i.tertiles ib2.grant Age BMI Years_educated i.addsaltattable
stepwise, pr(.1): logit Diabetes sex Exercise residence  tertiles grant Age BMI
Years_educated alcohol tobbaco addsaltattable
ereturn display, eform(or)
save "E:\PhD\Papers\Paper 2\Analytical data _ 7 Sept 20.dta"
////////// 7 September 2020
use "E:\PhD\Papers\Paper 2\Analytical data _ 7 Sept 20.dta"
cd "E:\PhD\Papers\Paper 2"
ren id ID
destring ID, gen(id)
```

```
merge 1:1 id using "E:\PhD\Papers\Paper 2\Prov codes.dta"
egen nmis = rmiss2(Diabetes currently_working Age_cat BMI sex Years_educated
Hypertension prov tertiles)

keep if (nmis == 0)
bugsdat (Age_cat BMI sex Years_educated tobbaco alcohol  Hypertension prov tertiles)
use "E:\PhD\Papers\Paper 2\Analytical data _ 7 Sept 20.dta"
replace tobbaco=3 if tobbaco==1
replace addsaltattable =3 if addsaltattable ==1
replace grant =3 if grant ==1
replace tobbaco=1 if tobbaco==2
replace tobbaco=2 if tobbaco==3
replace addsaltattable =2 if addsaltattable ==3
replace addsaltattable =1 if addsaltattable ==2
replace grant =2 if grant ==3
replace grant =1 if grant ==2
 egen nmis = rmiss2(Hypertension Age BMI sex Years_educated tobbaco Diabetes
addsaltattable district tertiles grant)
keep if (nmis == 0)
bugsdat (Hypertension Age BMI sex Years_educated tobbaco Diabetes addsaltattable district
tertiles grant)
```

Appendix D: Stata code for Paper 3

```
cd "E:\PhD\Papers\Paper 3"
use "E:\PhD\Papers\Paper 3\Analytical_no_stroke.dta", clear
tab1 q6001-q6009
tab1 q6001-q6009, nol
pca q6001-q6009
predict scindex
xtile scindex_cat = scindex , nq(3)
ta scindex_cat
recode q4001 (1 = 1 "Yes") (2 = 0 "No"), gen(arthritis)
recode q4014 (1 = 1 "Yes") (2 = 0 "No"), gen(angina)
recode q4022 (1 = 1 "Yes") (2 = 0 "No"), gen(diabetes)
recode q4025 (1 = 1 "Yes") (2 = 0 "No"), gen(lung)
recode q4033 (1 = 1 "Yes") (2 = 0 "No"), gen(asthma)
recode q4060 (1 = 1 "Yes") (2 = 0 "No"), gen(hypertension)
egen comorb1=rowtotal(arthritis angina diabetes lung asthma hypertension),
missing
recode comorb1 (0 1 = 0 "No") (2/9 = 1 "Yes"), gen(MM)
pca q0701 q0702 q0703 q0704 q0705 q0706 q0707 q0708 q0709 q0710 q0711
q0712 q0713 q0714 q0715 q0716 q0717 q0718 q0719 q0508 q0401 q0510
predict wealthindex
codebook wealthindex
*xtile quintiles = wealthindex , nq(5)
*ta quintiles
xtile tertiles = wealthindex , nq(3)
ta tertiles
/////Environmental safety: Q6018
recode BMI (0/18.49999999=1 "Underweight") (18.5/24.9999999=2 "Healthy
weight")  (25/29.999=3 "Overweight") (30/1000=4 "Obese"), gen(BMI_cat)

recode q6018 (4 5=1 "not safe") (3=2 "moderate") (1 2=3 "safe"),
gen(env_safety)
ren Exercise exercise1
gen Exercise=1 if q3016==1 | q3025==1 | q3019==1 | q3028==1
```

```
 replace Exercise=0 if Exercise!=1
 replace Exercise=. if q3028==. | q3025==. | q3016==. | q3019==.
 summtab, catvars(sex Ethnicity alcohol tobbaco addsaltattable Exercise
residence Sleep_quality scindex_cat env_safety tertiles) contvars(Age BMI WC
HC WHtR Years_educated) by(MM) word wordname(Table3_by_latentclasses)
median mnfmt(2) contptype(1) catptype(1) total pval title(Table 1:
Characteristics of participants by latent class category (n = 1967)) replace
 svyset q0101b [pw= adjpweight], strata(strata)
 tab MM [aw=adjpweight]
 *tab quintiles Classes [aw=adjpweight], row
 tab tertiles MM [aw=adjpweight], row
 tab sex [aw=adjpweight]
 tab Ethnicity [aw=adjpweight]
 svy: logit MM i.sex if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM ib2.alcohol if strata!=5 & strata!=12 & strata!=17 & strata!=18,
or
 svy: logit MM ib2.tobbaco if strata!=5 & strata!=12 & strata!=17 & strata!=18,
or
 svy: logit MM ib2.addsaltattable if strata!=5 & strata!=12 & strata!=17 &
strata!=18, or
 svy: logit MM i.Exercise if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM i.residence if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM i.tertiles if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM i.qosleep if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM i.scindex_cat if strata!=5 & strata!=12 & strata!=17 & strata!=18,
or
 svy: logit MM i.env_safety if strata!=5 & strata!=12 & strata!=17 & strata!=18,
or
 svy: logit MM Age if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM i.Ethnicity if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 svy: logit MM Years_educated if strata!=5 & strata!=12 & strata!=17 &
strata!=18, or
 ******Multivariable analyses
 svy: logit MM i.sex Age  i.Ethnicity i.tertiles Years_educated  i.scindex_cat
```

```
ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise  i.qosleep i.residence
i.env_safety   if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 replace Depression =0 if Depression==2
 svy: logit Depression i.MM i.sex Age  i.Ethnicity i.tertiles Years_educated
i.scindex_cat ib2.alcohol ib2.tobbaco ib2.addsaltattable i.Exercise  i.qosleep
i.residence i.env_safety if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 **stepwise, pr(.2): logit MM sex ib2.alcohol ib2.tobbaco ib2.addsaltattable
Exercise residence tertiles qosleep scindex_cat env_safety Age BMI
Years_educated if strata!=5 & strata!=12 & strata!=17 & strata!=18, or
 **tabulate env_safety, generate(ev_s)
 **tabulate qosleep, generate(qos)
 recode scindex_cat (1=0) (2 3=1), gen(SC)
 recode qosleep (2 3=0) (1=1), gen(SQ)
 recode tertiles (2 3=1) (1=0), gen(Wealth)
 recode env_safety (1=0) (2 3=1), gen(ES)
 replace tobbaco=tobbaco-1
 replace addsaltattable=addsaltattable-1
 replace alcohol=alcohol-1
 pca sex Age Ethnicity
 predict Demographic
 estat kmo
 pca Wealth Years_educated SC
 predict Socio_economic
 estat kmo
 pca SQ addsaltattable alcohol tobbaco Exercise
 predict Behavioural
 estat kmo
 pca ES residence
 predict Environmental
 estat kmo
 gsem (Socio_economic -> Depression, family(binomial) link(logit))
(Socio_economic -> MM, family(binomial) link(logit)) (Behavioural ->
Depression, family(binomial) link(logit)) (Behavioural -> MM, family(binomial)
link(logit)) (Environmental -> Depression, family(binomial) link(logit))
```

```
(Environmental -> MM, family(binomial) link(logit)) (Demographic ->
Depression, family(binomial) link(logit)) (Demographic -> MM,
family(binomial) link(logit)) (MM -> Depression, family(binomial) link(logit)),
nocapslatent
 ereturn display, eform(or)
```

Appendix E: Ethical clearance certificate

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Mr Glory Chidumwa et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M190919

NAME: (Principal Investigator)	Mr Glory Chidumwa et al
DEPARTMENT:	Epidemiology and Biostatistics School of Public Health
PROJECT TITLE:	Chronic non-communicable disease multimorbidities and syndemics in South African adults: Evidence from the WHO SAGE study using GSEM, MTBNs and Bayesian approaches
DATE CONSIDERED:	27/09/2019
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr L.J. Ware, Dr I. Maposa and Prof L. Micklesfield
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	24/10/2019

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

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **September** and will therefore be due in the month of **September** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 04-11-2019

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

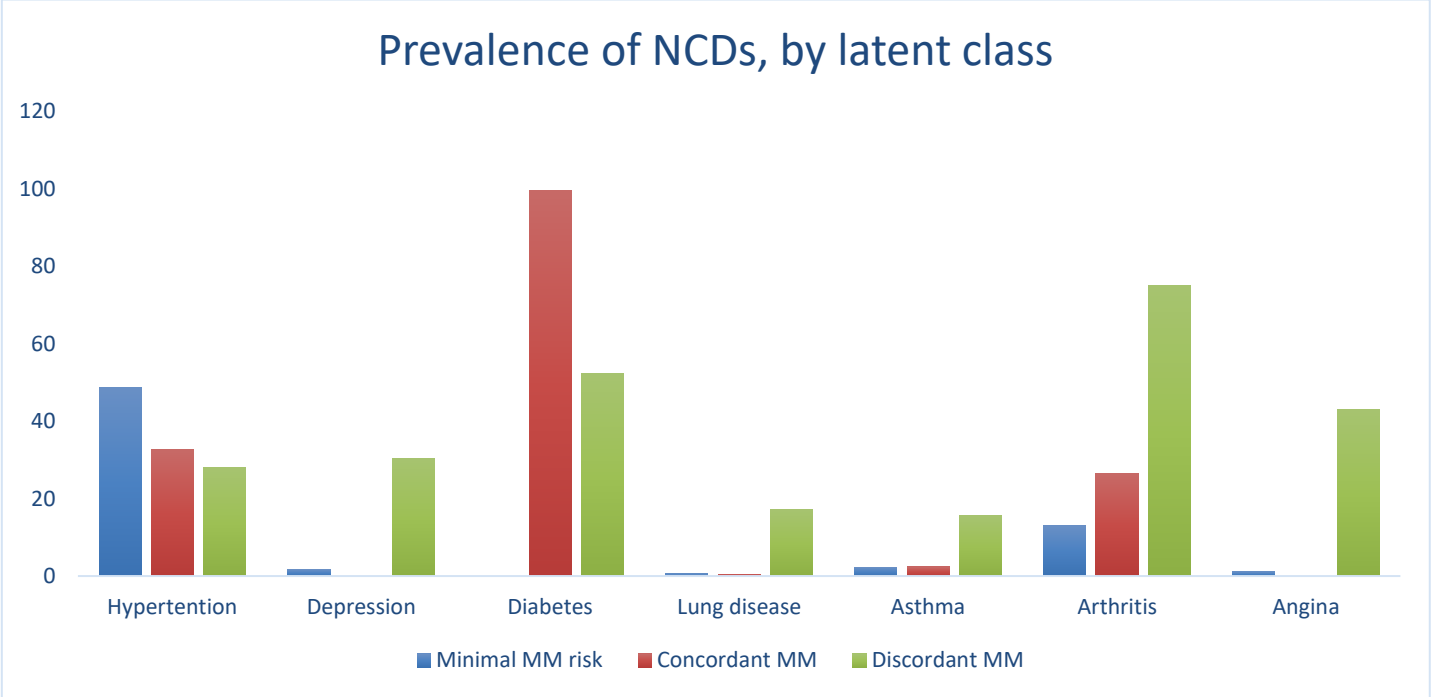
Appendix F: Supplementary tables and figures

Supplementary Table 3. 1: Latent class analysis fit statistics

Classes	G-squared	df	AIC	BIC	CAIC	aBIC	Entropy
2	357.29	985	433.29	658.37	696.37	537.63	0.70
3	302.14	965	418.14	761.69	819.69	577.41	0.71
4	267.79	945	423.79	885.81	963.81	637.98	0.74
5	242.56	925	438.56	1019.05	1117.05	707.67	0.78

AIC Aikake Information Criterion, BIC Bayesian Information Criterion

Supplementary Figure 3. 1: Prevalence of NCDs, by latent class



Supplementary Table 4. 1: Population size and sample size, by province

Province	Population estimate	% of total population	Number of 50+ sampled	% of 50+ sampled
	N=54,956,900		***N=1,817***	
Eastern Cape	6,916,200	*12.6*	287	*15.8*
Free State	2,817,900	*5.1*	157	*8.6*
Gauteng	13,200,300	*24.0*	321	*17.7*
KwaZulu-Natal	10,919,100	*19.9*	311	*17.1*
Limpopo	5,726,800	*10.4*	108	*5.9*
Mpumalanga	4,283,900	*7.8*	220	*12.1*
Northern Cape	1,185600	*2.2*	62	*3.4*
North West	3,707,000	*6.7*	116	*6.4*
Western Cape	6,200,100	*11.3*	235	*12.9*

Source: *Statistics SA 'Mid-year population estimates' (2015) Table 2 p 2.*

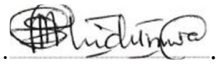
Supplementary Table 4. 2: Results of fitting stepwise backward elimination multiple logistic regression models for factors associated with hypertension and diabetes (p=0.10).

VARIABLE	Description	HYPERTENSION		DIABETES	
		aOR (95% CI)	P-value	aOR (95% CI)	P-value
Demographic characteristics					
Age (years)		1.05 (1.04; 1.06)	<0.001	1.04 (1.03; 1.05)	<0.001
Sex					
	Male	Reference		Reference	
	Female	2.08 (1.46; 2.98)	<0.001	1.86 (1.3; 2.65)	0.001
Socio-economic status characteristics					
Household wealth tertile					
	1 [lowest]	Reference		Reference	
	2	0.82 (0.57; 1.16)	0.257	0.59 (0.42; 0.85)	0.004
	3 [highest]	0.63 (0.41; 0.97)	0.037	0.42 (0.27; 0.64)	<0.001
Support from government?					
	No	Reference			
	Yes	1.37 (1.00; 1.87)	0.046		
Years of schooling		0.91 (0.87; 0.96)	<0.001	0.93 (0.88; 0.97)	0.002
Anthropological characteristics					
BMI		1.03 (1.01; 1.05)	0.004		
Behavioral characteristics					
Add salt at table?	*No*				
	Yes	1.85 (1.37; 2.50)	<0.001		
Ever used tobacco?	*No*			Reference	
	Yes			1.62 (1.12; 2.32)	0.010

CI- Confidence Interval; aOR- adjusted odds ratio; Results are from multivariate logistic regression models adjusting for all the potential covariates listed in Table 2. Only significant variables are reported.

Appendix G: Declaration: Student's contribution to articles and agreement of co-authors

I, **Glory Chidumwa**, student number; **1332997**, declare that this thesis is my own work and that I contributed adequately towards research findings published in the articles stated below which are included in my thesis.

Signature of Student……..... **Date:** ……31 July 2021.....…

Name of Primary Supervisor……Dr Lisa J Ware……

Signature of Primary Supervisor ……..**Date:** 08 December 2021

Agreement by co-authors: By signing this declaration, the co-authors listed below agree to the use of the article by the student as part of his/her Thesis/Dissertation/Research Report. In cases where the student is not the 1st author of a published article, the primary supervisor must explain (under comments) why the student is entitled to use the paper for his/her degree purposes.

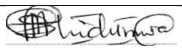
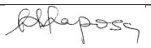
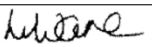
Article 1: Chidumwa G, Maposa I, Corso B, Minicuci N, Kowal P, Micklesfield LK, Ware LJ (2021) *Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa* Wave 2. BMJ open 11 (1):e041604

Authors	Name	Signature	Date
1st author	Glory Chidumwa		31 July 2021
2nd author	Innocent Maposa		30 August 2021
3rd author	Barbara Corso		9 December 2021
4th author	Nadia Minicuci		9 December 2021
5th author	Paul Kowal		09 Dec 2021
6th author	Lisa K Micklesfield		6 October 2021
7th author	Lisa J Ware		08 December 2021

Comments by primary supervisor:

This was a good manuscript from the student and he dealt well with the reviewers comments.

Article 2: Chidumwa G, Maposa I, Kowal P, Micklesfield LK, Ware LJ (2021) *Bivariate Joint Spatial Modeling to Identify Shared Risk Patterns of Hypertension and Diabetes in South Africa: Evidence from WHO SAGE South Africa Wave 2.* International Journal of Environmental Research and Public Health 18 (1):359

Authors	**Name**	**Signature**	**Date**
1st author	Glory Chidumwa		31 July 2021
2nd author	Innocent Maposa		30 August 2021
3rd author	Paul Kowal		09 Dec 2021
4th author	Lisa K Micklesfield		6 October 2021
5th author	Lisa J Ware		08 December 2021

Comments by primary supervisor:

This was a good manuscript from the student and he dealt well with the reviewers comments.

……

Appendix H: WHO-SAGE generic questionnaire

World Health Organization

STUDY ON GLOBAL AGEING AND ADULT HEALTH
Wave 1

Individual Questionnaire – Set A

sage

Table of Contents

Contact Record – Individual or Proxy Respondent 2
Section 1000: Socio-Demographic Characteristics 4
Section 1500: Work History and Benefits 8
Section 2000: Health State Descriptions 10
Section 2500: Anthropometrics, Performance Tests and Biomarkers 16
Section 3000: Risk Factors and Preventive Health Behaviours 26
Section 4000: Chronic Conditions and Health Services Coverage 30
Section 5000: Health Care Utilisation 38
Section 6000: Social Cohesion 50
Section 7000: Subjective Well-Being and Quality of Life 53
Section 8000: Impact of Caregiving 69
Section 9000: Interviewer Assessment 80