BMJ Open

Latent class analysis for chronic disease clusters: Evidence from SAGE South Africa Wave 2

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-041604
Article Type:	Original research
Date Submitted by the Author:	15-Jun-2020
Complete List of Authors:	Chidumwa, Glory; University of the Witwatersrand, Epidemiology and Biostatistics Maposa, Innocent; University of the Witwatersrand, Epidemiology and Biostatistics Corso, Barbara ; National Research Council, Neuroscience Institute Minicuci, Nadia; National Research Council, Neuroscience Institute Kowal, Paul ; Chiang Mai University Faculty of Science, Research Institute for Health Sciences Micklesfield, Lisa; University of the Witwatersrand, SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences Ware, Lisa; University of the Witwatersrand, SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences
Keywords:	Hypertension < CARDIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

TITLE

Latent class analysis for chronic disease clusters: Evidence from SAGE South Africa Wave 2

Authors

Glory Chidumwa[1], Innocent Maposa[1], Barbara Corso[2], Nadia Minicuci[2], Paul Kowal[3,4], Lisa K. Micklesfield[5], Lisa J Ware[5]

Affiliations:

1. Division of Epidemiology and Biostatistics, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa
2. Neuroscience Institute, National Research Council, Padova, Italy
3. World Health Organization (WHO), Geneva, Switzerland.
4. Research Institute for Health Sciences, Chiang Mai University, Chiang Mai, Thailand.
5. South African Medical Research Council/Wits Developmental Pathways for Health Research Unit, School of Clinical Medicine, University of the Witwatersrand, Johannesburg 2198, South Africa

Word count: 2812

Corresponding author:

Glory Chidumwa

Division of Epidemiology and Biostatistics

School of Public Health

Faculty of Health Sciences

University of the Witwatersrand

27 St Andrews Road

Parktown, 2193

Johannesburg

South Africa

E-mails: glorychidz@gmail.com; glory.chidumwa@wits.ac.za

Keywords

Multimorbidity; Latent Class Analysis; Chronic Non-communicable diseases; Multinomial logit

ABSTRACT

Objectives

Non-communicable diseases (NCDs) are the leading cause of global mortality and morbidity. In South Africa, NCDs were estimated to account for 57% of the total burden of disease in 2016. The aim of this study was to classify South African adults with chronic health conditions for multimorbidity risk, and to determine sociodemographic, anthropometric and behavioural factors associated with identified patterns of multimorbidity (MM), using data from the WHO Study on global AGEing and adult health (WHO SAGE) South Africa Wave 2.

Design

Cross-sectional study.

Participants

Data were retrieved from the WHO Study on Global AGEing and Adult Health Wave 2 for South Africa. A total of 1,967 individuals (men: 623, and women: 1,344) aged ≥45 years were included in the final analysis.

Measures

Multimorbidity latent classes.

Methods

Latent Class Analysis (LCA) was used on seven chronic conditions to identify multimorbidity latent classes. Multinomial logistic regression was used to determine which sociodemographic, anthropometric and behavioural factors were associated with the multimorbidity latent classes.

Results

The prevalence of multimorbidity (co-existence of two or more NCDs) was 21%. The LCA identified three groups namely: *minimal MM risk* (83%), *concordant (hypertension and diabetes) MM* (11%), and *discordant (angina, asthma, chronic lung disease, arthritis and depression) MM* (6%). Using the *minimal MM risk* group as the reference, female [RRR=4.57; 95% 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, while 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 high likelihood of belonging to the *discordant MM* group.

Conclusion

NCDs with similar pathophysiologic risk profiles tend to cluster together in older people. Risk factors for multimorbidity in South African adults include sex, age and tobacco use.

Strengths and limitations of this study

- This is the first comprehensive study on factors associated with the multimorbidity latent classes in low-income and middle-income countries.

- A key strength of the Study on global AGEing and adult health (SAGE) is that it consists of nationally representative samples, with high response rates.
- One weakness of this study is that 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.
- The cross-sectional design precludes causal inferences.

INTRODUCTION

Non-communicable diseases (NCDs) are the leading cause of mortality across the globe[1], and accounted for 73% of deaths in 2017[2, 3]. The prevalence of NCDs continues to increase in low- and middle-income countries (LMICs) including South Africa[1]. 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)[4-6]. 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[1, 7].

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[8, 9]. 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%[1]. 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[10]. 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[11]. 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[7]. 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[12].

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[13]. 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[13, 14]. 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 WHO 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.

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[15]. Further details are available on the WHO SAGE website (http://www.who.int/healthinfo/sage/en/). The current analysis includes 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)[16]. 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[17]. 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 - both leisure and work), and self-rated sleep quality (very good/good, moderate or poor/very poor) as reported previously[15]. 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[15]. 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[18-20].

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, such as regression models, as variation on observed indicators is modelled as a function of membership in unobserved classes called latent classes[21, 22]. 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[23]. 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[24]. 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. 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.

Ethics statement

This study used the WHO-SAGE Wave 1 data available in the public domain for use by researchers (http://www.who.int/healthinfo/sage/en/). The WHO-SAGE survey participants in all selected countries were informed about the survey, design, purpose and how it would benefit society at large. The survey was conducted under the supervision of the respective national governments. For this secondary data analysis, ethical clearance was obtained from the University of the Witwatersrand.

Patient and public involvement

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

RESULTS

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

Figure 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). The most common chronic disease was hypertension (52%) followed by arthritis (16%). Figure 2 below shows the prevalence of chronic NCDs by sex.

Figure 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[24, 25]. 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[13]. 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 demographic, anthropometric and behavioural characteristics of the three latent classes are presented in Table 1. The latent classes were significantly different with respect to all characteristics, with the exception of self-reported vigorous intensity activity. Details of the pairwise comparisons between the groups are shown in Table 1 below.

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

	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]	0.020
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	0.6 (0.5; 0.7)[a]	0.6 (0.6; 0.7)[b]	0.6 (0.6; 0.7)[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 2 below.

Table 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)	0.35
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)	0.540	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.

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[11, 12, 26, 27], 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[23].

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[28]. 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[28]. 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%[29]. 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)[29]. 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[25, 28, 29]. 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[25]. 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[26]. 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[30]. However, evidence suggests that implementation has faced challenges with many programmes remaining disease focused and with vertical implementation that fails to consider comorbid conditions[31, 32]. 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[33-35]. 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[36].

The findings from this study should be viewed in light of some limitations. Firstly, since the current study design is cross-sectional in nature, we could not determine the direction of the association 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. 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[27].

In conclusion, this study identified three latent classes namely: *minimal MM risk*, *concordant MM* and *discordant MM*. 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 strategies for prevention and intervention.

Summary table

What is known about the topic:

- MM clustering in Africa is generally assumed from research in high-income countries.

- The predictors of NCDs clustering and the management of NCDs in African populations require clarification.

What this study adds:

- MM prevalence remains high in South Africa
- Older individuals are more likely to be in one of the groups: minimal MM risk, concordant MM or discordant MM
- Risk factors for MM latent classes in South Africa include age, sex and tobacco use

Funding Sources:

WHO SAGE: Multi-country study is supported by 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 (OGHA 04034785; 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.

Authors' Contributions:

PK designed research; GC and BC performed analyses; GC and BC, LJW, IM, LKM, NM and PK wrote the paper; GC had primary responsibility for final content. All authors read and approved the final manuscript.

Conflicts of Interest:

The authors declare no conflict of interest.

Data sharing statement

The WHO SAGE data can be downloaded from the link:

https://www.who.int/healthinfo/sage/e. Data sharing statement

Ethics approval

SAGE received approval from the WHO's Ethical Review Committee and the respective committees in each participating country. Written informed consent was obtained from all study participants. For this secondary data analysis, ethical clearance was obtained from the University of the Witwatersrand.

Acknowledgments:

GC has had 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)'s 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 authors would also like to thank Dr Stephen Rule, Dr Robin Richards and Mr Godfrey Dlulane of Outsourced Insight who were subcontracted to conduct the surveys and coordinate data collection for WHO SAGE within South Africa. DPHRU acknowledge the support of the South African Medical Research Council. LJW is supported by the South African DSI-NRF Centre of Excellence in Human Development.

REFERENCES

1 Lalkhen H, Mash R. *Multimorbidity in non-communicable diseases in South African primary healthcare* 2015;**105**(2):134-138. DOI:10.7196/SAMJ.8696.

2 Vogeli C, Shields AE, Lee TA*, et al.* Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *Journal of general internal medicine* 2007;**22**:391-5.

3 Garin N, Koyanagi A, Olaya B*, et al.* Global Multimorbidity Patterns: A Cross-Sectional, Population-Based, Multi-Country Study. *The Journals of Gerontology: Series A* 2015;**71**:205-14.

4 Organization WH. *World health statistics 2010*: World Health Organization 2010.

5 Schouw D, Mash R, Kolbe-Alexander T. Transforming the workplace environment to prevent non-communicable chronic diseases: participatory action research in a South African power plant. *Global health action* 2018;**11**:1544336.

6 Ndinda C, Chilwane D, Mokomane Z. Civil society activism in accessing healthcare in South Africa: technical report. 2013.

7 Mayosi BM, Flisher AJ, Lalloo UG*, et al.* The burden of non-communicable diseases in South Africa. *Lancet (London, England)* 2009;**374**:934-47.

8 Holden L, Scuffham PA, Hilton MF*, et al.* Patterns of multimorbidity in working Australians. *Population health metrics* 2011;**9**:15.

9 Sciences AoM. Multimorbidity: a priority for global health research. Academy of Medical Sciences London 2018.

10 Afshar S, Roderick PJ, Kowal P*, et al.* Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. *BMC Public Health* 2015;**15**:776.

11 Ayeni OA, Norris SA, Joffe M*, et al.* The multimorbidity profile of South African women newly diagnosed with breast cancer. *Int J Cancer* 2019. DOI:10.1002/ijc.32727

12 Weimann A, Dai D, Oni T. 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* 2016;**163**:144-56.

13 Piette JD, Kerr EA. The Impact of Comorbid Chronic Conditions on Diabetes Care. *Diabetes Care* 2006;**29**:725-31.

14 Niankara I, Niankara A. 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* 2017;**2**:1-12.

15 Kowal P, Chatterji S, Naidoo N*, et al.* Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). *International journal of epidemiology* 2012;**41**:1639-49.

16 Ware LJ, Charlton K, Schutte AE*, et al.* Associations between dietary salt, potassium and blood pressure in South African adults: WHO SAGE Wave 2 Salt & Tobacco. *Nutr Metab Cardiovasc Dis* 2017;**27**:784-91.

17 Ware LJ, Chidumwa G, Charlton K*, et al.* Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2). *Journal of human hypertension* 2019;**33**:157-66.

18 Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy Plan* 2006;**21**:459-68.

19 Filmer D, Pritchett LH. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography* 2001;**38**:115-32.

20 McKenzie D. Measure inequality with asset indicators: Bureau for Research and Economic Analysis of Development. *Centre for International Development, BREAD* 2003.

21 Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 1974;**61**:215-31.

22 Land KC. Introduction to the special issue on finite mixture models. Sage Publications 2001.

23 Vermunt JK, Magidson J. Latent class cluster analysis. *Applied latent class analysis* 2002;**11**:89-106.

24 Nylund KL, Asparouhov T, Muthén BO. 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* 2007;**14**:535-69.

25 Olaya B, Moneta MV, Caballero FF*, et al.* Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. *BMC geriatrics* 2017;**17**:186.

26 Chang AY, Gómez-Olivé FX, Manne-Goehler J*, et al.* Multimorbidity and care for hypertension, diabetes and HIV among older adults in rural South Africa. *Bulletin of the World Health Organization* 2019;**97**:10.

27 Arokiasamy P, Uttamacharya U, Jain K*, et al.* 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* 2015;**13**:178.

28 Islam MM, Valderas JM, Yen L*, et al.* Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. *PloS one* 2014;**9**.

29 Whitson HE, Johnson KS, Sloane R*, et al.* Identifying patterns of multimorbidity in older Americans: application of latent class analysis. *Journal of the American Geriatrics Society* 2016;**64**:1668-73.

30 Mahomed OH, Asmall S, Freeman M. 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* 2014;**25**:1723-9.

31 Mahomed OH, Asmall S. Professional nurses' perceptions and experiences with the implementation of an integrated chronic care model at primary healthcare clinics in South Africa. *Curationis* 2017;**40**:1-6.

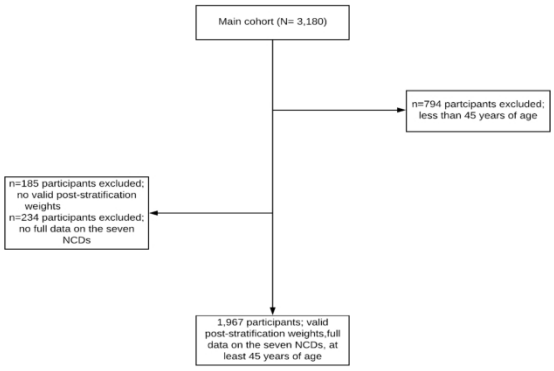
32 Mahomed OH, Asmall S, Voce A. Sustainability of the integrated chronic disease management model at primary care clinics in South Africa. *African journal of primary health care & family medicine* 2016;**8**:e1-e7.

33 Anand A, Sk MIK. The risk of hypertension and other chronic diseases: Comparing smokeless tobacco with smoking. *Frontiers in public health* 2017;**5**:255.

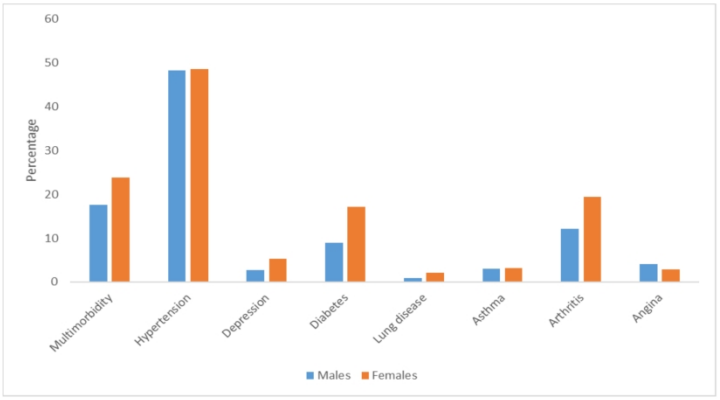
34 Boffetta P, Straif K. Use of smokeless tobacco and risk of myocardial infarction and stroke: systematic review with meta-analysis. *BMJ : British Medical Journal* 2009;**339**:b3060.

35 Thankappan K, Thresia C. Tobacco use & social status in Kerala. *Indian Journal of Medical Research* 2007;**126**:300.

36 Fonda JR, Gregor KL, Fortier CB*, et al.* 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* 2019;**236**:1729-39.



254x155mm (300 x 300 DPI)



254x150mm (300 x 300 DPI)

Reporting checklist for cross sectional study.

Based on the STROBE cross sectional guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE cross sectionalreporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title and abstract			
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1

Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	#3	State specific objectives, including any prespecified hypotheses	7
Methods			
Study design	#4	Present key elements of study design early in the paper	7
Setting	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of selection of participants.	7,8
	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	7

Bias	#9	Describe any efforts to address potential sources of bias	
Study size	#10	Explain how the study size was arrived at	7,8
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	8
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	9,10
Statistical methods	#12b	Describe any methods used to examine subgroups and interactions	
Statistical methods	#12c	Explain how missing data were addressed	
Statistical methods	#12d	If applicable, describe analytical methods taking account of sampling strategy	8
Statistical methods	#12e	Describe any sensitivity analyses	
Results			
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable.	11
Participants	#13b	Give reasons for non-participation at each stage	11

Participants	#13c	Consider use of a flow diagram	11
Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable.	11,12,13
Descriptive data	#14b	Indicate number of participants with missing data for each variable of interest	
Outcome data	#15	Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable.	13
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
Main results	#16b	Report category boundaries when continuous variables were categorized	
Main results	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	#18	Summarise key results with reference to study objectives	15

Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	18
Generalisability	#21	Discuss the generalisability (external validity) of the study results	
Other Information			
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19, 20

BMJ Open

Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-041604.R1
Article Type:	Original research
Date Submitted by the Author:	24-Oct-2020
Complete List of Authors:	Chidumwa, Glory; University of the Witwatersrand, Epidemiology and Biostatistics Maposa, Innocent; University of the Witwatersrand, Epidemiology and Biostatistics Corso, Barbara ; National Research Council, Neuroscience Institute Minicuci, Nadia; National Research Council, Neuroscience Institute Kowal, Paul ; Chiang Mai University Faculty of Science, Research Institute for Health Sciences Micklesfield, Lisa; University of the Witwatersrand, SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences Ware, Lisa; University of the Witwatersrand, SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology
Keywords:	Hypertension < CARDIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1

TITLE

Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2

Authors

Glory Chidumwa[1], Innocent Maposa[1], Barbara Corso[2], Nadia Minicuci[2], Paul Kowal[3,4], Lisa K. Micklesfield[5], Lisa J Ware[5]

Affiliations:

1. Division of Epidemiology and Biostatistics, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa
2. Neuroscience Institute, National Research Council, Padova, Italy
3. World Health Organization (WHO), Geneva, Switzerland.
4. Research Institute for Health Sciences, Chiang Mai University, Chiang Mai, Thailand.
5. South African Medical Research Council/Wits Developmental Pathways for Health Research Unit, School of Clinical Medicine, University of the Witwatersrand, Johannesburg 2198, South Africa

Word count: 2812

Corresponding author:

Glory Chidumwa

Division of Epidemiology and Biostatistics

School of Public Health

Faculty of Health Sciences

University of the Witwatersrand

27 St Andrews Road

Parktown, 2193

Johannesburg

South Africa

E-mails: glorychidz@gmail.com; glory.chidumwa@wits.ac.za

Keywords

Multimorbidity; Latent Class Analysis; Chronic Non-communicable diseases; Multinomial logit

ABSTRACT

Objectives

To classify South African adults with chronic health conditions for multimorbidity risk, and to determine sociodemographic, anthropometric and behavioural factors associated with identified patterns of multimorbidity (MM), using data from the World Health Organization Study on global AGEing and adult health (WHO SAGE) South Africa Wave 2.

Design

Nationally representative (for ≥50 years old adults) cross-sectional study.

Setting: Adults in South Africa between 2014 and 2015.

Participants

1,967 individuals (men: 623, and women: 1,344) aged ≥45 years for whom data on all 7 health conditions and socioeconomic, demographic, behavioral, and anthropological information were available.

Measures

Multimorbidity latent classes.

Methods

Latent Class Analysis (LCA) was used on seven chronic conditions to identify multimorbidity latent classes. Multinomial logistic regression was used to determine which sociodemographic, anthropometric and behavioural factors were associated with the multimorbidity latent classes.

Results

The prevalence of multimorbidity (co-existence of two or more non-communicable diseases (NCDs)) was 21%. The LCA identified three groups namely: *minimal MM risk* (83%), *concordant (hypertension and diabetes) MM* (11%), and *discordant (angina, asthma, chronic lung disease, arthritis and depression) MM* (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, while 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 high likelihood of belonging to the *discordant MM* group.

Conclusion

NCDs with similar pathophysiologic risk profiles tend to cluster together in older people. Risk factors for multimorbidity in South African adults include sex, age and tobacco use.

Strengths and limitations of this study

- This is the first comprehensive study on factors associated with the multimorbidity latent classes in low-income and middle-income countries.
- A key strength of the Study on global AGEing and adult health (SAGE) is that it consists of nationally representative samples, with high response rates.
- One weakness of this study is that 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.

- The cross-sectional design precludes causal inferences.

INTRODUCTION

Non-communicable diseases (NCDs) are the leading cause of mortality across the globe[1], and accounted for 73% of deaths in 2017[2, 3]. 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[4, 5]. The prevalence of NCDs continues to increase in low- and middle-income countries (LMICs) including South Africa[1]. 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)[6-8]. 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[1, 9].

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[10, 11]. 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%[1]. 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%) [3]. 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[12]. 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[13]. 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[9]. 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[14].

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[15]. 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[15, 16]. 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.

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[17]. 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)[18]. 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[19]. 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[17]. 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[17]. 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[20-22].

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[23, 24]. 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[25]. Furthermore, LCA has been demonstrated to be more objective and rigorous than K-means and hierarchical clustering for both exploratory work and theory testing [26]. 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[25]. 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[27]. 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.

Ethics statement

This study used the WHO-SAGE Wave 2 data available in the public domain for use by researchers (http://www.who.int/healthinfo/sage/en/). The WHO-SAGE survey participants in all selected countries were informed about the survey, design, purpose and how it would benefit society at large. The survey was conducted under the supervision of the respective national governments. For this secondary data analysis, ethical clearance was obtained from the University of the Witwatersrand.

Patient and public involvement

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

RESULTS

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

Figure 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 2 below shows the prevalence of chronic NCDs by sex.

Figure 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[27, 28]. 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[15]. 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 1.

The demographic, anthropometric and behavioural characteristics of the three latent classes are presented in Table 1. The latent classes were significantly different with respect to all characteristics, with the exception of self-reported vigorous intensity activity. Details of the pairwise comparisons between the groups are shown in Table 1 below.

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

	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]	0.020
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 2 below.

Table 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)	0.35
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)	0.540	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.

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[13, 14, 29, 30], 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[25].

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[31]. 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[31]. 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%[32]. 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)[32]. 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[28, 31, 32]. 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[28]. 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[29]. 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[33]. However, evidence suggests that implementation has faced challenges with many programmes remaining disease focused and with vertical implementation that fails to consider comorbid conditions[34, 35]. 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[36-38]. 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[39].

The findings from this study should be viewed in light of some limitations. Firstly, since the current study design is cross-sectional in nature, we could not determine the direction of the association 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. 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[30].

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 in 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 policy makers to prioritize the older population, females and tobacco users in prevention programs.

Funding Sources:

WHO SAGE: Multi-country study is supported by 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 (OGHA 04034785; 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.

Authors' Contributions:

PK designed research; GC and BC performed analyses; GC and BC, LJW, IM, LKM, NM and PK wrote the paper; GC had primary responsibility for final content. All authors read and approved the final manuscript.

Conflicts of Interest:

The authors declare no conflict of interest.

Data sharing statement

The WHO SAGE data can be downloaded from the link:

https://www.who.int/healthinfo/sage/e. Data sharing statement

Ethics approval

SAGE received approval from the WHO's Ethical Review Committee and the respective committees in each participating country. Written informed consent was obtained from all study participants. For this secondary data analysis, ethical clearance was obtained from the University of the Witwatersrand.

Acknowledgments:

GC has had 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)'s 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 authors would also like to thank Dr Stephen Rule, Dr Robin Richards and Mr Godfrey Dlulane of Outsourced Insight who were subcontracted to conduct the surveys and coordinate data collection for WHO SAGE within South Africa. DPHRU acknowledge the support of the South African Medical Research Council. LJW is supported by the South African DSI-NRF Centre of Excellence in Human Development.

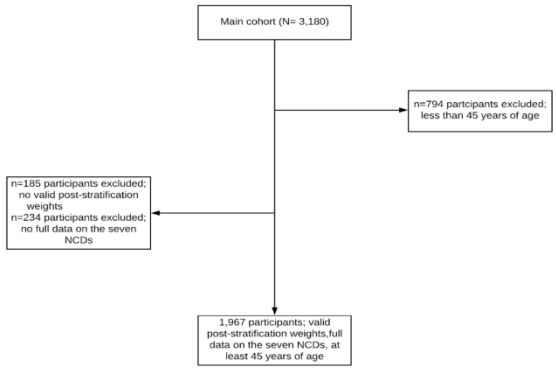
REFERENCES

1. Lalkhen H, Mash R. Multimorbidity in non-communicable diseases in South African primary healthcare2015.
2. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. Journal of general internal medicine. 2007;22(3):391-5.
3. Garin N, Koyanagi A, Olaya B, Lara E, Haro JM, Tyrovolas S, et al. Global Multimorbidity Patterns: A Cross-Sectional, Population-Based, Multi-Country Study. The Journals of Gerontology: Series A. 2015;71(2):205-14.
4. Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L. Prevalence of multimorbidity among adults seen in family practice. The Annals of Family Medicine. 2005;3(3):223-8.
5. Britt HC, Harrison CM, Miller GC, Knox SA. Prevalence and patterns of multimorbidity in Australia. Medical Journal of Australia. 2008;189(2):72-7.
6. Organization WH. World health statistics 2010: World Health Organization; 2010.
7. Schouw D, Mash R, Kolbe-Alexander T. Transforming the workplace environment to prevent non-communicable chronic diseases: participatory action research in a South African power plant. Global health action. 2018;11(1):1544336.
8. Ndinda C, Chilwane D, Mokomane Z. Civil society activism in accessing healthcare in South Africa: technical report. 2013.
9. Mayosi BM, Flisher AJ, Lalloo UG, Sitas F, Tollman SM, Bradshaw D. The burden of non-communicable diseases in South Africa. Lancet (London, England). 2009;374(9693):934-47.
10. Holden L, Scuffham PA, Hilton MF, Muspratt A, Ng S-K, Whiteford HA. Patterns of multimorbidity in working Australians. Population health metrics. 2011;9(1):15.
11. Sciences AoM. Multimorbidity: a priority for global health research. Academy of Medical Sciences London; 2018.
12. Afshar S, Roderick PJ, Kowal P, Dimitrov BD, Hill AG. Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. BMC Public Health. 2015;15(1):776.
13. Ayeni OA, Norris SA, Joffe M, Cubasch H, Nietz S, Buccimazza I, et al. The multimorbidity profile of South African women newly diagnosed with breast cancer. Int J Cancer. 2019.
14. Weimann A, Dai D, Oni T. 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. 2016;163:144-56.
15. Piette JD, Kerr EA. The Impact of Comorbid Chronic Conditions on Diabetes Care. Diabetes Care. 2006;29(3):725-31.
16. Niankara I, Niankara A. 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. 2017;2(1):1-12.
17. Kowal P, Chatterji S, Naidoo N, Biritwum R, Fan W, Lopez Ridaura R, et al. Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). International journal of epidemiology. 2012;41(6):1639-49.
18. Ware LJ, Charlton K, Schutte AE, Cockeran M, Naidoo N, Kowal P. Associations between dietary salt, potassium and blood pressure in South African adults: WHO SAGE Wave 2 Salt & Tobacco. Nutr Metab Cardiovasc Dis. 2017;27(9):784-91.
19. Ware LJ, Chidumwa G, Charlton K, Schutte AE, Kowal P. Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2). Journal of human hypertension. 2019;33(2):157-66.
20. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan. 2006;21(6):459-68.

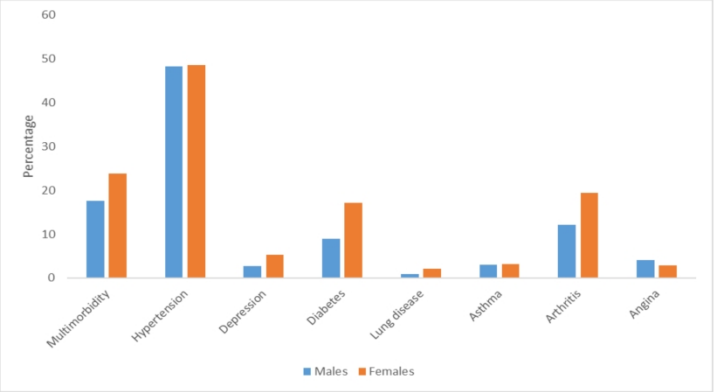
21. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography. 2001;38(1):115-32.
22. McKenzie D. Measure inequality with asset indicators: Bureau for Research and Economic Analysis of Development. Centre for International Development, BREAD. 2003.
23. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika. 1974;61(2):215-31.
24. Land KC. Introduction to the special issue on finite mixture models. Sage Publications; 2001.
25. Vermunt JK, Magidson J. Latent class cluster analysis. Applied latent class analysis. 2002;11:89-106.
26. Schreiber JB, Pekarik AJ. Using Latent Class Analysis versus K-means or Hierarchical Clustering to understand museum visitors. Curator: The Museum Journal. 2014;57(1):45-59.
27. Nylund KL, Asparouhov T, Muthén BO. 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. 2007;14(4):535-69.
28. Olaya B, Moneta MV, Caballero FF, Tyrovolas S, Bayes I, Ayuso-Mateos JL, et al. Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. BMC geriatrics. 2017;17(1):186.
29. Chang AY, Gómez-Olivé FX, Manne-Goehler J, Wade AN, Tollman S, Gaziano TA, et al. Multimorbidity and care for hypertension, diabetes and HIV among older adults in rural South Africa. Bulletin of the World Health Organization. 2019;97(1):10.
30. Arokiasamy P, Uttamacharya U, Jain K, Biritwum RB, Yawson AE, Wu F, et al. 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. 2015;13(1):178.
31. Islam MM, Valderas JM, Yen L, Dawda P, Jowsey T, McRae IS. Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. PloS one. 2014;9(1).
32. Whitson HE, Johnson KS, Sloane R, Cigolle CT, Pieper CF, Landerman L, et al. Identifying patterns of multimorbidity in older Americans: application of latent class analysis. Journal of the American Geriatrics Society. 2016;64(8):1668-73.
33. Mahomed OH, Asmall S, Freeman M. 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. 2014;25(4):1723-9.
34. Mahomed OH, Asmall S. Professional nurses' perceptions and experiences with the implementation of an integrated chronic care model at primary healthcare clinics in South Africa. Curationis. 2017;40:1-6.
35. Mahomed OH, Asmall S, Voce A. Sustainability of the integrated chronic disease management model at primary care clinics in South Africa. African journal of primary health care & family medicine. 2016;8(1):e1-e7.
36. Anand A, Sk MIK. The risk of hypertension and other chronic diseases: Comparing smokeless tobacco with smoking. Frontiers in public health. 2017;5:255.
37. Boffetta P, Straif K. Use of smokeless tobacco and risk of myocardial infarction and stroke: systematic review with meta-analysis. BMJ : British Medical Journal. 2009;339:b3060.
38. Thankappan K, Thresia C. Tobacco use & social status in Kerala. Indian Journal of Medical Research. 2007;126(4):300.
39. Fonda JR, Gregor KL, Fortier CB, Scioli-Salter ER, McGlinchey RE, Rasmusson A. 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. 2019;236(6):1729-39.

Supplementary Table 1: Latent class analysis fit statistics

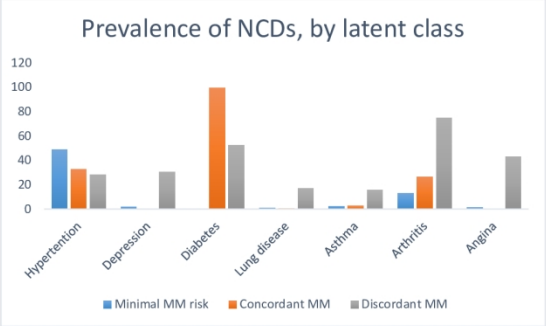
Supplementary Figure 1: Prevalence of NCDs, by latent class



254x155mm (300 x 300 DPI)



254x150mm (300 x 300 DPI)



Supplementary Figure 1: Prevalence of NCDs, by latent class

215x279mm (200 x 200 DPI)

Supplementary Table 1: Latent class analysis fit statistics

Classes	G-squared	DF	AIC	BIC	CAIC	**aBIC**	**Entropy**
2	228.14	477	296.14	497.53	531.53	**389.50**	**0.70**
3	**191.89**	**459**	**295.89**	**603.90**	**655.90**	**438.68**	**0.83**
4	154.58	441	294.58	709.22	779.22	**486.80**	**0.62**
5	145.35	423	321.35	841.93	930.60	**563.00**	**0.53**

AIC-Akaike Information Criterion, BIC-Bayesian Information Criterion, DF-degrees of freedoms, aBIC-sample size adjusted BIC

Reporting checklist for cross sectional study.

Based on the STROBE cross sectional guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE cross sectionalreporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title and abstract			
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1

Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	#3	State specific objectives, including any prespecified hypotheses	7
Methods			
Study design	#4	Present key elements of study design early in the paper	7
Setting	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of selection of participants.	7,8
	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	7

Bias	#9	Describe any efforts to address potential sources of bias	
Study size	#10	Explain how the study size was arrived at	7,8
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	8
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	9,10
Statistical methods	#12b	Describe any methods used to examine subgroups and interactions	
Statistical methods	#12c	Explain how missing data were addressed	
Statistical methods	#12d	If applicable, describe analytical methods taking account of sampling strategy	8
Statistical methods	#12e	Describe any sensitivity analyses	
Results			
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable.	11
Participants	#13b	Give reasons for non-participation at each stage	11

Participants	#13c	Consider use of a flow diagram	11
Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable.	11,12,13
Descriptive data	#14b	Indicate number of participants with missing data for each variable of interest	
Outcome data	#15	Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable.	13
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
Main results	#16b	Report category boundaries when continuous variables were categorized	
Main results	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	#18	Summarise key results with reference to study objectives	15

Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	18
Generalisability	#21	Discuss the generalisability (external validity) of the study results	
Other Information			
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19, 20

BMJ Open

Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-041604.R2
Article Type:	Original research
Date Submitted by the Author:	25-Nov-2020
Complete List of Authors:	Chidumwa, Glory; University of the Witwatersrand, Epidemiology and Biostatistics Maposa, Innocent; University of the Witwatersrand, Epidemiology and Biostatistics Corso, Barbara ; National Research Council, Neuroscience Institute Minicuci, Nadia; National Research Council, Neuroscience Institute Kowal, Paul ; Chiang Mai University Faculty of Science, Research Institute for Health Sciences Micklesfield, Lisa; University of the Witwatersrand, SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences Ware, Lisa; University of the Witwatersrand, SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology
Keywords:	Hypertension < CARDIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1

TITLE

Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2

Authors

Glory Chidumwa[1], Innocent Maposa[1], Barbara Corso[2], Nadia Minicuci[2], Paul Kowal[3,4], Lisa K. Micklesfield[5], Lisa J Ware[5]

Affiliations:

1. Division of Epidemiology and Biostatistics, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa
2. Neuroscience Institute, National Research Council, Padova, Italy
3. World Health Organization (WHO), Geneva, Switzerland.
4. Research Institute for Health Sciences, Chiang Mai University, Chiang Mai, Thailand.
5. South African Medical Research Council/Wits Developmental Pathways for Health Research Unit, School of Clinical Medicine, University of the Witwatersrand, Johannesburg 2198, South Africa

Word count: 2812

Corresponding author:

Glory Chidumwa

Division of Epidemiology and Biostatistics

School of Public Health

Faculty of Health Sciences

University of the Witwatersrand

27 St Andrews Road

Parktown, 2193

Johannesburg

South Africa

E-mails: glorychidz@gmail.com; glory.chidumwa@wits.ac.za

Keywords

Multimorbidity; Latent Class Analysis; Chronic Non-communicable diseases; Multinomial logit

ABSTRACT

Objectives

To classify South African adults with chronic health conditions for multimorbidity risk, and to determine sociodemographic, anthropometric and behavioural factors associated with identified patterns of multimorbidity (MM), using data from the World Health Organization Study on global AGEing and adult health (WHO SAGE) South Africa Wave 2.

Design

Nationally representative (for ≥50 years old adults) cross-sectional study.

Setting: Adults in South Africa between 2014 and 2015.

Participants

1,967 individuals (men: 623, and women: 1,344) aged ≥45 years for whom data on all 7 health conditions and socioeconomic, demographic, behavioral, and anthropological information were available.

Measures

Multimorbidity latent classes.

Results

The prevalence of multimorbidity (co-existence of two or more non-communicable diseases (NCDs)) was 21%. The LCA identified three groups namely: *minimal MM risk* (83%), *concordant (hypertension and diabetes) MM* (11%), and *discordant (angina, asthma, chronic lung disease,*

arthritis and depression) MM (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, while 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 high likelihood of belonging to the *discordant MM* group.

Conclusion

NCDs with similar pathophysiologic risk profiles tend to cluster together in older people. Risk factors for multimorbidity in South African adults include sex, age and tobacco use.

Strengths and limitations of this study

- This is the first comprehensive study on factors associated with the multimorbidity latent classes in low-income and middle-income countries.
- A key strength of the Study on global AGEing and adult health (SAGE) is that it consists of nationally representative samples, with high response rates.
- One weakness of this study is that 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.
- The cross-sectional design precludes causal inferences.

INTRODUCTION

Non-communicable diseases (NCDs) are the leading cause of mortality across the globe[1], and accounted for 73% of deaths in 2017[2, 3]. 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[4, 5]. The prevalence of NCDs continues to increase in low- and middle-income countries (LMICs) including South Africa[1]. 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)[6-8]. 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[1, 9].

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[10, 11]. 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%[1]. 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%) [3]. 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[12]. 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[13]. 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[9]. 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[14].

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[15]. 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[15, 16]. 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.

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[17]. 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)[18]. 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[19]. 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[17]. 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[17]. 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[20-22].

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[23, 24]. 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[25]. Furthermore, LCA has been demonstrated to be more objective and rigorous than K-means and hierarchical clustering for both exploratory work and theory testing [26]. 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[25]. 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[27]. 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.

Ethics statement

This study used the WHO-SAGE Wave 2 data available in the public domain for use by researchers (http://www.who.int/healthinfo/sage/en/). The WHO-SAGE survey participants in all selected countries were informed about the survey, design, purpose and how it would benefit society at large. The survey was conducted under the supervision of the respective national governments. For this secondary data analysis, ethical clearance was obtained from the University of the Witwatersrand.

Patient and public involvement

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

RESULTS

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

Figure 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 2 below shows the prevalence of chronic NCDs by sex.

Figure 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[27, 28]. 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[15]. 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 1.

The demographic, anthropometric and behavioural characteristics of the three latent classes are presented in Table 1. The latent classes were significantly different with respect to all characteristics, with the exception of self-reported vigorous intensity activity. Details of the pairwise comparisons between the groups are shown in Table 1 below.

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

	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]	0.020
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 2 below.

Table 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)	0.35
	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)	0.540	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.

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[13, 14, 29, 30], 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[25].

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[31]. 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[31]. 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%[32]. 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)[32]. 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[28, 31, 32]. 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[28]. 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[29]. 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[33]. However, evidence suggests that implementation has faced challenges with many programmes remaining disease focused and with vertical implementation that fails to consider comorbid conditions[34, 35]. 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[36-38]. 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[39].

The findings from this study should be viewed in light of some limitations. Firstly, since the current study design is cross-sectional in nature, we could not determine the direction of the association 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[30].

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 in 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 policy makers to prioritize the older population, females and tobacco users in prevention programs.

Funding Sources:

WHO SAGE: Multi-country study is supported by 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 (OGHA 04034785; 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.

Authors' Contributions:

PK designed research; GC and BC performed analyses; GC and BC, LJW, IM, LKM, NM and PK wrote the paper; GC had primary responsibility for final content. All authors read and approved the final manuscript.

Conflicts of Interest:

The authors declare no conflict of interest.

Data sharing statement

The WHO SAGE data can be downloaded from the link:

https://www.who.int/healthinfo/sage/e. Data sharing statement

Ethics approval

SAGE received approval from the WHO's Ethical Review Committee and the respective committees in each participating country. Written informed consent was obtained from all study participants. For this secondary data analysis, ethical clearance was obtained from the University of the Witwatersrand.

Acknowledgments:

GC has had 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)'s 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 authors would also like to thank Dr Stephen Rule, Dr Robin Richards and Mr Godfrey Dlulane of Outsourced Insight who were subcontracted to conduct the surveys and coordinate data collection for WHO SAGE within

South Africa. DPHRU acknowledge the support of the South African Medical Research Council.

LJW is supported by the South African DSI-NRF Centre of Excellence in Human Development.

REFERENCES

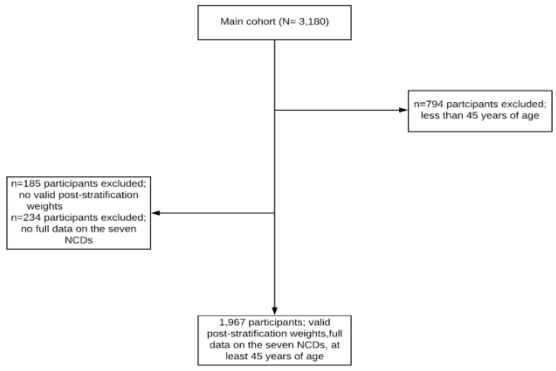
1. Lalkhen H, Mash R. Multimorbidity in non-communicable diseases in South African primary healthcare2015.
2. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. Journal of general internal medicine. 2007;22(3):391-5.
3. Garin N, Koyanagi A, Olaya B, Lara E, Haro JM, Tyrovolas S, et al. Global Multimorbidity Patterns: A Cross-Sectional, Population-Based, Multi-Country Study. The Journals of Gerontology: Series A. 2015;71(2):205-14.
4. Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L. Prevalence of multimorbidity among adults seen in family practice. The Annals of Family Medicine. 2005;3(3):223-8.
5. Britt HC, Harrison CM, Miller GC, Knox SA. Prevalence and patterns of multimorbidity in Australia. Medical Journal of Australia. 2008;189(2):72-7.
6. Organization WH. World health statistics 2010: World Health Organization; 2010.
7. Schouw D, Mash R, Kolbe-Alexander T. Transforming the workplace environment to prevent non-communicable chronic diseases: participatory action research in a South African power plant. Global health action. 2018;11(1):1544336.
8. Ndinda C, Chilwane D, Mokomane Z. Civil society activism in accessing healthcare in South Africa: technical report. 2013.
9. Mayosi BM, Flisher AJ, Lalloo UG, Sitas F, Tollman SM, Bradshaw D. The burden of non-communicable diseases in South Africa. Lancet (London, England). 2009;374(9693):934-47.
10. Holden L, Scuffham PA, Hilton MF, Muspratt A, Ng S-K, Whiteford HA. Patterns of multimorbidity in working Australians. Population health metrics. 2011;9(1):15.
11. Sciences AoM. Multimorbidity: a priority for global health research. Academy of Medical Sciences London; 2018.
12. Afshar S, Roderick PJ, Kowal P, Dimitrov BD, Hill AG. Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. BMC Public Health. 2015;15(1):776.
13. Ayeni OA, Norris SA, Joffe M, Cubasch H, Nietz S, Buccimazza I, et al. The multimorbidity profile of South African women newly diagnosed with breast cancer. Int J Cancer. 2019.
14. Weimann A, Dai D, Oni T. 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. 2016;163:144-56.
15. Piette JD, Kerr EA. The Impact of Comorbid Chronic Conditions on Diabetes Care. Diabetes Care. 2006;29(3):725-31.
16. Niankara I, Niankara A. 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. 2017;2(1):1-12.
17. Kowal P, Chatterji S, Naidoo N, Biritwum R, Fan W, Lopez Ridaura R, et al. Data resource profile: the World Health Organization Study on global AGEing and adult health (SAGE). International journal of epidemiology. 2012;41(6):1639-49.

18. Ware LJ, Charlton K, Schutte AE, Cockeran M, Naidoo N, Kowal P. Associations between dietary salt, potassium and blood pressure in South African adults: WHO SAGE Wave 2 Salt & Tobacco. Nutr Metab Cardiovasc Dis. 2017;27(9):784-91.
19. Ware LJ, Chidumwa G, Charlton K, Schutte AE, Kowal P. Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2). Journal of human hypertension. 2019;33(2):157-66.
20. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan. 2006;21(6):459-68.
21. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography. 2001;38(1):115-32.
22. McKenzie D. Measure inequality with asset indicators: Bureau for Research and Economic Analysis of Development. Centre for International Development, BREAD. 2003.
23. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika. 1974;61(2):215-31.
24. Land KC. Introduction to the special issue on finite mixture models. Sage Publications; 2001.
25. Vermunt JK, Magidson J. Latent class cluster analysis. Applied latent class analysis. 2002;11:89-106.
26. Schreiber JB, Pekarik AJ. Using Latent Class Analysis versus K-means or Hierarchical Clustering to understand museum visitors. Curator: The Museum Journal. 2014;57(1):45-59.
27. Nylund KL, Asparouhov T, Muthén BO. 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. 2007;14(4):535-69.
28. Olaya B, Moneta MV, Caballero FF, Tyrovolas S, Bayes I, Ayuso-Mateos JL, et al. Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. BMC geriatrics. 2017;17(1):186.
29. Chang AY, Gómez-Olivé FX, Manne-Goehler J, Wade AN, Tollman S, Gaziano TA, et al. Multimorbidity and care for hypertension, diabetes and HIV among older adults in rural South Africa. Bulletin of the World Health Organization. 2019;97(1):10.
30. Arokiasamy P, Uttamacharya U, Jain K, Biritwum RB, Yawson AE, Wu F, et al. 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. 2015;13(1):178.
31. Islam MM, Valderas JM, Yen L, Dawda P, Jowsey T, McRae IS. Multimorbidity and comorbidity of chronic diseases among the senior Australians: prevalence and patterns. PloS one. 2014;9(1).
32. Whitson HE, Johnson KS, Sloane R, Cigolle CT, Pieper CF, Landerman L, et al. Identifying patterns of multimorbidity in older Americans: application of latent class analysis. Journal of the American Geriatrics Society. 2016;64(8):1668-73.
33. Mahomed OH, Asmall S, Freeman M. 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. 2014;25(4):1723-9.
34. Mahomed OH, Asmall S. Professional nurses' perceptions and experiences with the implementation of an integrated chronic care model at primary healthcare clinics in South Africa. Curationis. 2017;40:1-6.
35. Mahomed OH, Asmall S, Voce A. Sustainability of the integrated chronic disease management model at primary care clinics in South Africa. African journal of primary health care & family medicine. 2016;8(1):e1-e7.
36. Anand A, Sk MIK. The risk of hypertension and other chronic diseases: Comparing smokeless tobacco with smoking. Frontiers in public health. 2017;5:255.

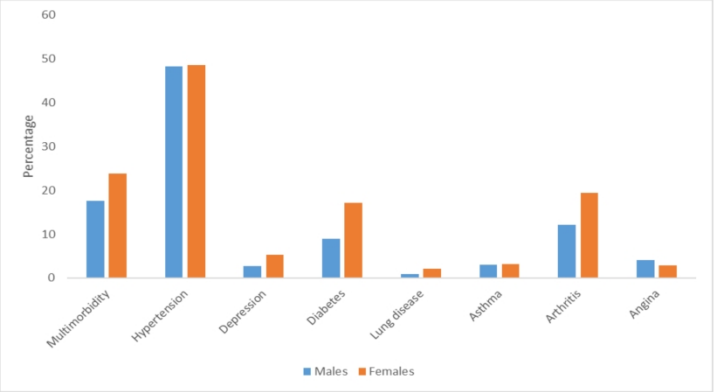
37. Boffetta P, Straif K. Use of smokeless tobacco and risk of myocardial infarction and stroke: systematic review with meta-analysis. BMJ : British Medical Journal. 2009;339:b3060.
38. Thankappan K, Thresia C. Tobacco use & social status in Kerala. Indian Journal of Medical Research. 2007;126(4):300.
39. Fonda JR, Gregor KL, Fortier CB, Scioli-Salter ER, McGlinchey RE, Rasmusson A. 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. 2019;236(6):1729-39.

Supplementary Table 1: Latent class analysis fit statistics

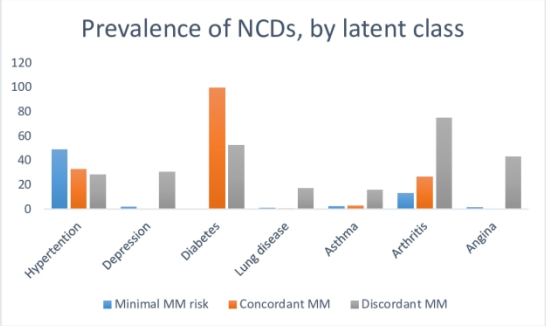
Supplementary Figure 1: Prevalence of NCDs, by latent class



254x155mm (300 x 300 DPI)



254x150mm (300 x 300 DPI)



Supplementary Figure 1: Prevalence of NCDs, by latent class

215x279mm (200 x 200 DPI)

Supplementary Table 1: Latent class analysis fit statistics

Classes	G-squared	DF	AIC	BIC	CAIC	**aBIC**	**Entropy**
2	228.14	477	296.14	497.53	531.53	**389.50**	**0.70**
3	**191.89**	**459**	**295.89**	**603.90**	**655.90**	**438.68**	**0.83**
4	154.58	441	294.58	709.22	779.22	**486.80**	**0.62**
5	145.35	423	321.35	841.93	930.60	**563.00**	**0.53**

AIC-Akaike Information Criterion, BIC-Bayesian Information Criterion, DF-degrees of freedoms, aBIC-sample size adjusted BIC

Reporting checklist for cross sectional study.

Based on the STROBE cross sectional guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE cross sectionalreporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title and abstract			
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1

Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	#3	State specific objectives, including any prespecified hypotheses	7
Methods			
Study design	#4	Present key elements of study design early in the paper	7
Setting	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of selection of participants.	7,8
	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	7

Bias	#9	Describe any efforts to address potential sources of bias	
Study size	#10	Explain how the study size was arrived at	7,8
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	8
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	9,10
Statistical methods	#12b	Describe any methods used to examine subgroups and interactions	
Statistical methods	#12c	Explain how missing data were addressed	
Statistical methods	#12d	If applicable, describe analytical methods taking account of sampling strategy	8
Statistical methods	#12e	Describe any sensitivity analyses	
Results			
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable.	11
Participants	#13b	Give reasons for non-participation at each stage	11

Participants	#13c	Consider use of a flow diagram	11
Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable.	11,12,13
Descriptive data	#14b	Indicate number of participants with missing data for each variable of interest	
Outcome data	#15	Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable.	13
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	14
Main results	#16b	Report category boundaries when continuous variables were categorized	
Main results	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	#18	Summarise key results with reference to study objectives	15

Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	18
Generalisability	#21	Discuss the generalisability (external validity) of the study results	
Other Information			
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19, 20