


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

Glory Chidumwa <sup>1</sup>, Innocent Maposa,<sup>1</sup> Barbara Corso,<sup>2</sup> Nadia Minicuci,<sup>2</sup> Paul Kowal,<sup>3</sup> Lisa K Micklesfield,<sup>4</sup> Lisa Jayne Ware <sup>4,5</sup>

**To cite:** Chidumwa G, Maposa I, Corso B, *et al*. Identifying co-occurrence and clustering of chronic diseases using latent class analysis: cross-sectional findings from SAGE South Africa Wave 2. *BMJ Open* 2021;**11**:e041604. doi:10.1136/bmjopen-2020-041604

► Prepublication history and supplemental material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-041604>).

Received 15 June 2020  
Revised 25 November 2020  
Accepted 08 January 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Glory Chidumwa;  
[glory.chidumwa@wits.ac.za](mailto:glory.chidumwa@wits.ac.za)

## ABSTRACT

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

**Design** Nationally representative (for ≥50-year-old adults) cross-sectional study.

**Setting** Adults in South Africa between 2014 and 2015.

**Participants** 1967 individuals (men: 623 and women: 1344) aged ≥45 years for whom data on all seven health conditions and socioeconomic, demographic, behavioural, and anthropological information were available.

**Measures** MM latent classes.

**Results** The prevalence of MM (coexistence of two or more non-communicable diseases (NCDs)) was 21%. The latent class analysis 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% CI (1.64 to 12.75); p=0.004) and older (RRR=1.08; 95% CI (1.04 to 1.12); p<0.001) participants were more likely to belong to the concordant MM group, while tobacco users (RRR=8.41; 95% CI (1.93 to 36.69); p=0.005) and older (RRR=1.09; 95% CI (1.03 to 1.15); p=0.002) participants had a high likelihood of belonging to the discordant MM group.

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

## INTRODUCTION

Non-communicable diseases (NCDs) are the leading cause of mortality across the globe,<sup>1</sup> and accounted for 73% of deaths in 2017.<sup>2,3</sup> In developed countries, it is estimated that approximately one in every four adults experience multimorbidity (MM), with half of older adults having three or more chronic conditions.<sup>4,5</sup> The prevalence of NCDs continues to increase in low/middle-income countries

## 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 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), were based on self-report, and can thus be affected by possible recall bias and social desirability bias.
- The cross-sectional design precludes causal inferences.

(LMICs) including South Africa.<sup>1</sup> 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).<sup>6–8</sup> 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.<sup>1,9</sup>

Within an individual, the coexistence of two or more chronic non-communicable, mental health or infectious diseases, of long duration (>3 months), is referred to as MM.<sup>10,11</sup> Data from a 2015 South African primary healthcare (PHC) survey across all age groups reported the prevalence of NCD MM, which included hypertension, diabetes, ischaemic heart disease, asthma, epilepsy, chronic obstructive pulmonary disease, osteoarthritis and respiratory infection, as 14.4%.<sup>1</sup> A study by Garin *et al* aimed at identifying and describing MM patterns among adults older than 50 years in low-income, middle-income and high-income countries, using data from the Collaborative Research on Ageing in Europe project and the WHO's Study on global AGEing and adult health (WHO-SAGE) Wave 1, found

that South Africa had a higher prevalence (68%) of MM (having at least two NCDs) than Ghana (48%), India (58%) and China (45%).<sup>3</sup> In addition, in a study by Afshar *et al* to compare the prevalence of MM across 28 LMICs using the World Health Survey (2003), the prevalence of MM (two chronic conditions or more) in South Africa was 21.6% among the 50–64 year age group and 30.1% among those aged 65 years and older.<sup>12</sup> A study by Ayeni *et al* aimed at profiling MM among 2281 South African women of age 18 years and older, newly diagnosed with breast cancer, across two South African provinces.<sup>13</sup> They reported that 43.9% of the women met the definition of MM 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.<sup>9</sup> A study by Weimann *et al* investigated the association between socio-economic disadvantage and MM in South Africa at two time points, 2008 and 2012, using the National Income Dynamics Study. They showed that the risk for MM was doubled in urban residents relative to their rural counterparts, and respondents who were socioeconomically deprived had a twofold increased risk of having MM compared with the less deprived in both urban and rural areas.<sup>14</sup>

Previous research on MM in South Africa has primarily used simple counts of chronic conditions. However, different combinations of diseases may affect a person's health and healthcare differently.<sup>15</sup> To account for these differences, disease combinations can be categorised 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.<sup>15 16</sup> The prevalence and patterns of MM 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 MM risk, using self-reported diagnosed NCD health condition variables in a latent class analysis (LCA) using data from the WHO-SAGE South Africa Wave 2. Additionally, the analyses looked at sociodemographic, anthropometric and behavioural factors associated with identified patterns of MM. The findings of the current study will contribute to the evidence base on the epidemiology of MM 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 multicountry longitudinal study including China, Ghana, India, Mexico and the Russian Federation, to examine the health and well-being of nationally representative adult populations aged

18+ years in over 42 000 participants, with an emphasis on populations aged 50+ years.<sup>17</sup> Further details are available on the WHO-SAGE website (<http://www.who.int/health-info/sage/en/>). The current study is a cross-sectional analysis for the SAGE South Africa Wave 2 data collected in 2014/2015 using participants (n=1967), 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 categorised 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).<sup>18</sup> Hypertension status was determined as a measured average systolic blood pressure reading of  $\geq 140$  mm Hg; and/or an average diastolic blood pressure reading of  $\geq 90$  mm Hg; and/or current use (within the last 2 weeks) of antihypertensive medication.<sup>19</sup> 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 min 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 min continuously?'), and self-rated sleep quality (very good/good, moderate or poor/very poor) as reported previously.<sup>17</sup> 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.<sup>17</sup> Body mass index (weight, kg/height, m<sup>2</sup>) and waist-to-height ratio (waist (cm)/height (cm)) were calculated. Principal component 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.<sup>20–22</sup>

### Statistical analysis

Data were captured using an electronic data capture system (CAPI). STATA Statistical Software: V.16.0 (Stata Corp, 2017; College Station, USA) was used for statistical analyses. The LCA was performed in SAS (Version 9.4) PROC LCA add-on to determine patterns of coexisting 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.<sup>23 24</sup> 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.<sup>25</sup> Furthermore, LCA has been demonstrated to be more objective and rigorous than K-means and hierarchical clustering for both exploratory work and theory testing.<sup>26</sup> This is because LCA is model based, that is, there is a statistical model that is assumed to come from the population from which the data were gathered.<sup>25</sup> 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),

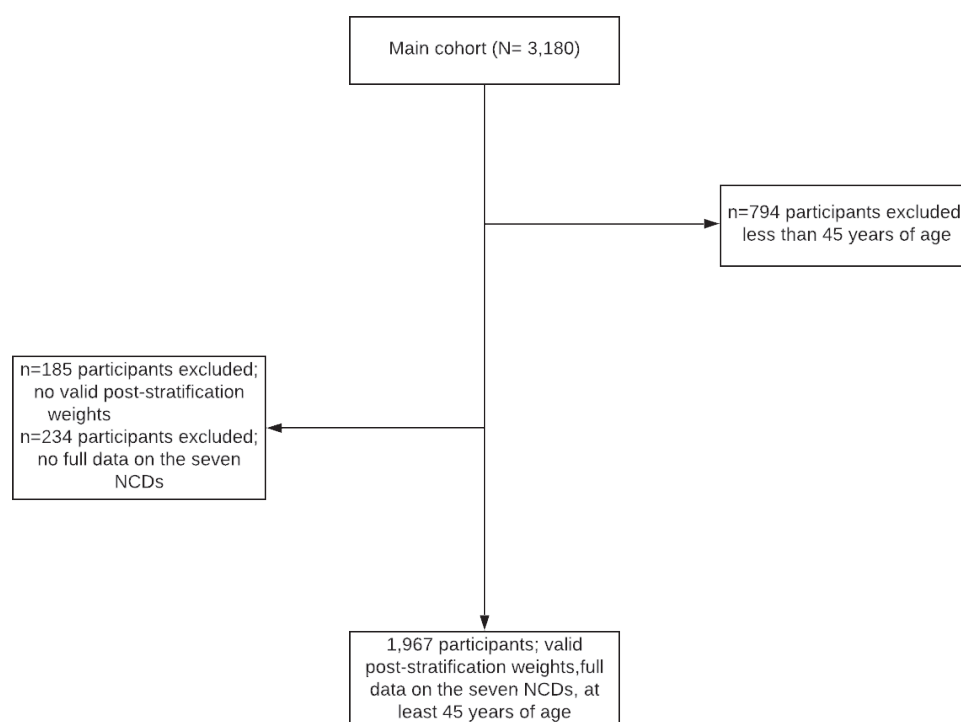
which has been shown to provide robust indicators of class enumeration with categorical outcomes.<sup>27</sup> The aBIC was used to compare several plausible class models where the lowest values indicate the best fitting model. After selecting the best model, each participant was assigned to one class according to his or her highest computed probability of membership. Details for the LCA fit statistics are given in online supplemental table 1. The Pearson's  $X^2$  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, 95% CIs and p values are reported for each explanatory variable.

### Ethics statement

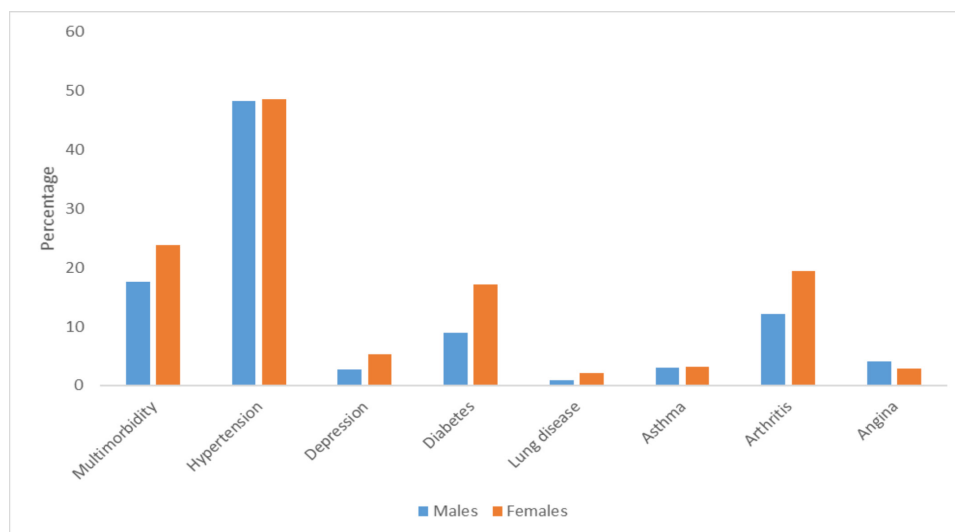
This study used the WHO-SAGE Wave 2 data available in the public domain for use by researchers (<http://www.who.int/healthinfo/sage/en/>). The WHO-SAGE survey participants in all selected countries were informed about the survey, design, purpose and how it would benefit society at large. The survey was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

### Patient and public involvement

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



**Figure 1** Study flow diagram. NCDs, non-communicable diseases.



**Figure 2** Prevalence of chronic NCDs by sex. NCDs, non-communicable diseases.

## RESULTS

A total of 1967 participants were included in this analysis. **Figure 1** below shows the study flow diagram.

The median age for the sample was 62 years (IQR: 54–70). Fifty-seven per cent (n=1113) of our sample were women. The majority of the sample self-identified as black (n=1540, 78%), 6% (n=120) as white, and 16% (n=308) as coloured or Indian.

### Prevalence of chronic NCDs and MM

Twenty-one per cent of the sample (n=415) had two or more of the seven chronic diseases, that is, 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.

The prevalence of arthritis, depression, diabetes, lung disease and MM was higher in women, and of angina was higher in men.

### Latent classes for chronic disease clusters

The optimal number of latent classes was determined using the aBIC. There were negligible differences between the two class and three class models and considering plausible interpretability, the three-class model was chosen.<sup>27 28</sup> 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 Kerr as chronic conditions that represent the similar pathophysiological 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.<sup>15</sup> The majority of the sample (n=1625, 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 is presented by latent class as online supplemental 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.

Multinomial logistic regression results showing associations between the demographic, anthropometric and behavioural characteristics, and latent class membership, are presented in **table 2** below.

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 1-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 was significant in this logistic regression.

## DISCUSSION

In this study, we have shown that the prevalence of MM (coexistence of two or more NCDs) was 21%. The LCA

**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) <sup>a</sup>	65 (58–72) <sup>b</sup>	62 (55.5–69) <sup>a</sup>	<0.001
BMI	28.5 (24.2–34.4) <sup>a</sup>	29.5 (25.6–35.6)	31.1 (25.2–37.5) <sup>b</sup>	0.02
Waist circumference (cm)	94 (81–105) <sup>a</sup>	99 (88–109) <sup>b</sup>	100 (88–112) <sup>b</sup>	<0.001
Hip circumference (cm)	100 (90–112) <sup>a</sup>	106 (94–116) <sup>b</sup>	106.5 (93–118) <sup>b</sup>	<0.001
Waist-to-height ratio	0.579 (0.503–0.662) <sup>a</sup>	0.642 (0.571–0.728) <sup>b</sup>	0.634 (0.553–0.710) <sup>b</sup>	<0.001
Years educated	8 (6–11) <sup>a</sup>	8 (5–10) <sup>b</sup>	8 (6–10)	0.023
Sex				<0.001
Male	545 (34.3) <sup>a</sup>	51 (20.6) <sup>b</sup>	27 (21.1) <sup>b</sup>	
Female	1046 (65.7)	197 (79.4)	101 (78.9)	
Alcohol				0.033
Yes	289 (18.2) <sup>a</sup>	31 (12.7) <sup>b</sup>	29 (22.8) <sup>a</sup>	
No	1296 (81.8)	214 (87.3)	98 (77.2)	
Tobacco				<0.001
Yes	301 (19.0) <sup>a</sup>	35 (14.3) <sup>a</sup>	40 (31.7)	
No	1284 (81.0)	210 (85.7)	86 (68.3)	
Add salt at table				0.013
Yes	1084 (68.4)	155 (63.3)	73 (57.0)	
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)	160 (64.5)	101 (78.9)	
Rural	467 (29.4)	88 (35.5)	27 (21.1)	
Household wealth tertile				0.001
1 (lowest)	395 (80.58)	39 (8.03)	56 (11.39)	
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)	176 (73.0)	89 (71.2)	
Bad	263 (16.8)	65 (27.0)	36 (28.8)	

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  $X^2$  test for categorical data; p values shown are for Kruskal-Wallis test for continuous data and pairwise  $X^2$  test for categorical data. For categorical data, frequencies are reported with percentages in parenthesis, while medians are reported for continuous data with IQRs in parenthesis.

BMI, body mass index; MM, multimorbidity.

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 with 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 MM in South Africa,<sup>13 14 29 30</sup> however this study has used data from

the SAGE which represents the 50+ years South African population, to identify patterns of chronic disease coexistence. In addition, to our knowledge this is the first study in South Africa to use LCA to identify patterns of chronic disease coexistence as LCA has the ability to identify unique combinations of diseases using probabilities.<sup>25</sup>

Our study identified three latent classes of MM based on the presence or absence of seven chronic conditions. Previous studies that have used the LCA method to describe patterns of chronic disease coexistence in

**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 to 1.12)	<0.001	1.09 (1.04 to 1.14)	0.001
Sex				
Male	Reference		Reference	
Female	4.38 (1.42 to 13.6)	0.011	2.04 (0.58 to 7.24)	0.267
Alcohol				
No	Reference		Reference	
Yes	1.13 (0.13 to 9.76)	0.908	0.37 (0.08 to 1.70)	0.201
Tobacco				
No	Reference		Reference	
Yes	2.92 (0.61 to 13.9)	0.178	8.86 (2.03 to 38.8)	0.004
Add salt at table				
No	Reference		Reference	
Yes	1.00 (0.43 to 2.33)	0.992	0.53 (0.23 to 1.22)	0.136
Physical activity				
No	Reference		Reference	
Yes	1.12 (0.48 to 2.61)	0.784	0.77 (0.26 to 2.30)	0.639
Residence				
Urban	Reference		Reference	
Rural	1.14 (0.41 to 3.21)	0.799	1.31 (0.43 to 4.00)	0.633
Household wealth tertile				
1 (lowest)	Reference		Reference	
2	1.10 (0.45 to 2.71)	0.833	0.61 (0.23 to 1.58)	0.303
3 (highest)	1.49 (0.38 to 5.8)	0.564	0.43 (0.05 to 3.75)	0.443
Sleep quality				
Good/very good	Reference		Reference	
Moderate	1.58 (0.67 to 3.72)	0.292	1.65 (0.57 to 4.77)	0.35
Poor/very poor	2.38 (0.66 to 8.55)	0.183	0.99 (0.23 to 4.34)	0.989
BMI	0.98 (0.92 to 1.04)	0.540	1.01 (0.97 to 1.06)	0.564
Years educated	1.00 (0.91 to 1.11)	0.861	1.01 (0.83 to 1.23)	0.939

BMI, body mass index; MM, multimorbidity.

older populations have yielded mixed results as regards the number of clusters identified. A cross-sectional sample of 4574 Australian senior citizens (aged 50 years and over) using 11 chronic conditions reported an MM prevalence of 52% and identified four classes.<sup>31</sup> Their sample presented (1) a relatively healthier group, (2) a sick group with dominant presence of arthritis, asthma and depression, (3) a sick group with dominant presence of hypertension and diabetes, and (4) the sickest group with dominant presence of cancer, heart and stroke.<sup>31</sup> Similarly, a retrospective cohort study on 13 self-reported conditions from 14502 Americans (65 years old and older) identified six classes using the LCA approach, and reported an MM of 67.3%.<sup>32</sup> The classes included: minimal disease class (prevalence of all conditions is below cohort average), non-vascular 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).<sup>32</sup> 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.<sup>28 31 32</sup> 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 *et al* 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.<sup>28</sup> 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 *et al* in rural South Africa where they defined concordant conditions as cardiometabolic conditions (hypertension, diabetes and angina), and discordant conditions as mental health illness, alcohol dependence and HIV infection.<sup>29</sup> 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 healthcare.<sup>33</sup> However, evidence suggests that implementation has faced challenges with many programmes remaining disease focused and with vertical implementation that fails to consider comorbid conditions.<sup>34 35</sup> 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 with non-tobacco users.<sup>36–38</sup> For example, in a study by Fonda *et al* aimed at examining the clustering of post-traumatic stress disorder, depressive disorders and clinically significant pain among 433 deployed veterans in Boston (USA) aged 18–65 years, tobacco smokers had 3.5 increased likelihood for MM.<sup>39</sup>

The findings from this study should be viewed in light of some limitations. First, 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), were 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 criteria 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 MM prevalence. However, our prevalence data for MM are similar overall to previous SAGE recent data, and a number of studies have also analysed MM 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.<sup>30</sup>

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 PHC system under the ICDM model remains single-disease focused on the treatment of patients. In improving PHC in South Africa, efforts should be made to manage multiple conditions concurrently at PHC centres, in particular diabetes and hypertension. In addition, in our sample, risk factors for MM latent classes include age, sex and tobacco use. Future efforts should focus on the inclusion of all frequently occurring common conditions, including infectious diseases to evaluate clustering patterns and inform policymakers to prioritise the older population, women and tobacco users in prevention programmes.

#### Author affiliations

<sup>1</sup>Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, Gauteng, South Africa

<sup>2</sup>Neuroscience Institute, National Research Council, Padova, Italy

<sup>3</sup>Research Institute for Health Sciences, Chiang Mai University Faculty of Science, Chiang Mai, Thailand

<sup>4</sup>SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg-Braamfontein, Gauteng, South Africa

<sup>5</sup>DSI-NRF Centre of Excellence in Human Development, University of the Witwatersrand, Johannesburg, Gauteng, South Africa

**Acknowledgements** GC has had support from the Developing Excellence in Leadership, Training and Science (DELTA) Africa Initiative. The DELTA Africa Initiative is an independent funding scheme of the African Academy of Sciences’ (AAS) Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa’s Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (grant 107754/Z/15/Z-DELTA 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 Dzulane 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.

**Contributors** PK designed the research. GC and BC performed analyses. GC, 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.

**Funding** The WHO-SAGE multicountry study is supported by the WHO and the Division of Behavioral and Social Research (BSR) at the National Institute on Aging (NIA), US National Institutes of Health, through Interagency Agreements (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.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**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.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. Data are available upon reasonable request. The WHO Study on global AGEing and adult health (SAGE) data are available upon request at: <https://www.who.int/healthinfo/sage/en/>.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Glory Chidumwa <http://orcid.org/0000-0002-8743-9045>

Lisa Jayne Ware <http://orcid.org/0000-0002-9762-4017>

#### REFERENCES

- Lalkhen H, Mash R. Multimorbidity in non-communicable diseases in South African primary healthcare. *S Afr Med J* 2015;105:134.
- Vogeli C, Shields AE, Lee TA, *et al*. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *J Gen Intern Med* 2007;22:391–5.
- Garin N, Koyanagi A, Chatterji S, Olaya B, *et al*. Global multimorbidity patterns: a cross-sectional, population-based, Multi-Country study. *J Gerontol A Biol Sci Med Sci* 2016;71:205–14.
- Fortin M, Bravo G, Hudon C, *et al*. Prevalence of multimorbidity among adults seen in family practice. *Ann Fam Med* 2005;3:223–8.
- Britt HC, Harrison CM, Miller GC, *et al*. Prevalence and patterns of multimorbidity in Australia. *Med J Aust* 2008;189:72–7.
- Organization WH. World health statistics 2010: World Health organization 2010.
- 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. *Glob Health Action* 2018;11:1544336.
- Ndinda C, Chilwane D, Mokomane Z. Civil society activism in accessing healthcare in South Africa: technical report 2013.
- Mayosi BM, Flisher AJ, Lalloo UG, *et al*. The burden of non-communicable diseases in South Africa. *The Lancet* 2009;374:934–47.
- Holden L, Scuffham PA, Hilton MF, *et al*. Patterns of multimorbidity in working Australians. *Popul Health Metr* 2011;9:15.
- AoM S. Multimorbidity: a priority for global health research. *Academy of Medical Sciences London* 2018.
- 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.
- Ayeni OA, Norris SA, Joffe M, *et al*. The multimorbidity profile of South African women newly diagnosed with breast cancer. *Int J Cancer* 2019.
- 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. *Soc Sci Med* 2016;163:144–56.
- Piette JD, Kerr EA. The impact of comorbid chronic conditions on diabetes care. *Diabetes Care* 2006;29:725–31.
- 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. *Clin Med Ther* 2017;2:1–12.
- Kowal P, Chatterji S, Naidoo N, *et al*. Data resource profile: the world Health organization study on global ageing and adult health (SAGE). *Int J Epidemiol* 2012;41:1639–49.
- 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.
- 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). *J Hum Hypertens* 2019;33:157–66.
- Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy Plan* 2006;21:459–68.
- 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.
- McKenzie D. Measure inequality with asset indicators: bureau for research and economic analysis of development. *Centre for International Development, BREAD* 2003.
- Goodman LEOA. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 1974;61:215–31.
- Land KC. *Introduction to the special issue on finite mixture models*. Sage Publications, 2001: 29. 275–81.
- Vermunt JK, Magidson J. Latent class cluster analysis. *Applied latent class analysis* 2002;11:89–106.
- Schreiber JB, Pekarik AJ. Technical note: using latent class analysis versus k-means or hierarchical clustering to understand museum visitors. *Curator: The Museum Journal* 2014;57:45–59.
- 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.
- 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 Geriatr* 2017;17:186.
- 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. *Bull World Health Organ* 2019;97:10–23.
- 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 Med* 2015;13:178.
- 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:e83783.
- Whitson HE, Johnson KS, Sloane R, *et al*. Identifying patterns of multimorbidity in older Americans: application of latent class analysis. *J Am Geriatr Soc* 2016;64:1668–73.
- Mahomed OH, Asmall S, Freeman M. An integrated chronic disease management model: a diagonal approach to health system strengthening in South Africa. *J Health Care Poor Underserved* 2014;25:1723–9.
- 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.
- Mahomed OH, Asmall S, Voce A. Sustainability of the integrated chronic disease management model at primary care clinics in South Africa. *Afr. j. prim. health care fam. med.* 2016;8:e1–7.
- Anand A, Sk MIK. The risk of hypertension and other chronic diseases: comparing smokeless tobacco with smoking. *Front Public Health* 2017;5:255.
- Boffetta P, Straif K. Use of smokeless tobacco and risk of myocardial infarction and stroke: systematic review with meta-analysis. *BMJ* 2009;339:b3060.
- Thankappan KR, Thresia CU. Tobacco use & social status in Kerala. *Indian J Med Res* 2007;126:300.
- 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.



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	<b>389.50</b>	<b>0.70</b>
<b>3</b>	<b>191.89</b>	<b>459</b>	<b>295.89</b>	<b>603.90</b>	<b>655.90</b>	<b>438.68</b>	<b>0.83</b>
4	154.58	441	294.58	709.22	779.22	<b>486.80</b>	<b>0.62</b>
5	145.35	423	321.35	841.93	930.60	<b>563.00</b>	<b>0.53</b>

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

