Extended Abstract

Title: Variation and Prevalence of Multimorbidity Clusters across Rural Africa, 2018-2021

Authors: Cyril Chironda, Chodziwadziwa Kabudula, Farai Mlambo, F Gomez-Olive Casas, Rumbidzai Mupfuti, Tobias Chirwa

Introduction

Multimorbidity, defined as the coexistence of two or more chronic conditions within an individual, poses significant challenges to global health systems, particularly in resource-limited settings of rural Africa. This study focuses on the prevalence and patterns of multimorbidity clusters among the deceased across several Health and Demographic Surveillance System (HDSS) sites in Africa from 2018 to 2021. The HDSS sites included in this research are ARHI, AGINCOURT and DIMAMO in South Africa, KARONGA in Malawi, HARAR and KERSA in Ethiopia, NAVRONGO in Ghana, and SIAYA in Kenya.

Aim and Objectives

This study aims to understand the variations in multimorbidity reported deaths across different African rural HDSS sites between 2018 and 2021. With the objective to understand the differences and commonalities in the prevalence of chronic conditions and their related clusters among the deceased. Furthermore, the study seeks to identify demographic and behavioural risk factors contributing to the resultant clusters of chronic conditions for each site.

Theoretical Focus

The theoretical framework for this study is grounded in the biopsychosocial model, which integrates biological, psychological, and social factors in disease manifestation and progression. The study hypothesises that multimorbidity cluster patterns vary significantly across different demographic and behavioural contexts.

Methods

Study Design

The study uses a quantitative longitudinal design, relying on retrospective secondary data from Verbal Autopsy (VA) collected over time in the HDSS sites, from 2018 to 2021.

Data

The study uses the VA data from, HDSS sites including ARHI, AGINCOURT and DIMAMO in South Africa, KARONGA in Malawi, HARAR and KERSA in Ethiopia, NAVRONGO in Ghana, and SIAYA in Kenya.VA is a method that is well-suited for routine application because it is utilised to identify causes of death in places without extensive medical certification systems. Trained fieldworkers conduct VA interviews with the deceased's family members or caregivers at the HDSS sites. Details regarding the signs, symptoms, circumstances, and opinions of the respondents regarding the causes of death are gathered during these interviews related to the terminal illness.

The study seeks to cluster the following chronic conditions: Tuberculosis, Human Immunodeficiency Virus, Hypertension, Heart Disease, Diabetes Mellitus, Asthma, Epilepsy, Cancer, Chronic Obstructive Pulmonary Disease, Dementia, Kidney Disease, Liver Disease. The

demographic and behavioural variables include: Marriage, Age, Gender, Alcohol use and Tobacco use.

The dataset contains 93,043 individuals from the eight sites, from 2018 to 2021. The distribution of participants varies significantly among the sites, with AHRI having the highest participation at 24,700 individuals, constituting 26.55% of the total, and Harar the lowest at 2,007 individuals, making up 2.16%. Notably, Siaya HDSS also contributes a significant portion with 16,424 individuals, representing 17.65% of the dataset. Other sites such as Agincourt and Navrongo contribute 20,373 (21.90%) and 12,448 (13.38%) individuals, respectively.

AHRI reports a nearly balanced gender distribution with 49.69% female and 50.31% male participants. In contrast, DIMAMO shows a significant gender disparity with 65.21% female and 34.79% male participants. Other sites such as NAVRONGO and Siaya HDSS also display varying gender proportions, with NAVRONGO having 42.42% female and 57.58% male, and Siaya HDSS showing a more balanced distribution with 46.81% female and 53.19% male.

Agincourt shows a substantial segment of its population in the 65 and older category, comprising 28.61%, while AHRI has 22.70% in the same age group, indicating a heavier skew towards older populations in these areas. Meanwhile, sites like Karonga and Kersa report higher percentages of younger age groups, with Karonga documenting 11.62% under one month old.

Analysis Methods

To achieve the research objective, we will utilize the VA data from the HDSS sites. Initially, we will calculate the counts and proportions of deceased individuals by gender and age group for each year of death (YOD) and site. By categorizing these individuals by age group and sex and then computing the counts and proportions for each year and site for every chronic condition, we will compile general statistics for the various sites under study.

Next, we will assess the distribution of comorbidities among deceased individuals over time. We will determine the number and proportion of people with 1, 2, and 3 or more comorbidities for each year and site. This analysis will serve as the initial step in identifying variations in the general prevalence of multimorbidities affecting populations across the different study sites. To visualize the counts of different combinations of comorbidities for each site, we will generate upset plots. These plots will display the raw combinations of comorbidities for each site. We will employ these groupings as supervisory variables to test and compare the efficacy of various clustering algorithms.

We will then cluster chronic conditions based on similarities in their prevalence and co-occurrence. Hierarchical clustering will be used to analyse the similarity matrix derived from co-occurrence data and to create a hierarchy of clusters. These clusters will be visualized using dendrograms, and we will examine how they differ across each site.

Furthermore, we will cluster deceased individuals based on their chronic condition profiles using k-medoids clustering. Each individual will be represented by their chronic condition profile, and we will analyse how these clusters vary for each site.

Latent class analysis will be implemented to identify latent classes of chronic condition profiles among deceased individuals. This involves running an algorithm to partition individuals into classes or clusters based on their chronic condition patterns, thereby maximizing the probability of the observed data given the latent class structure. We will explore how these classes differ for each site.

We will compare the clusters obtained from variable clustering, individual clustering, and latent class analysis with the groupings obtained from the upset plots. This comparison will inform us which clustering and classifying method performs better for each study setting.

Finally, we will conduct a multinomial regression analysis to identify factors contributing to the resultant clusters of chronic conditions for each site. The regression model is given by:

$$\text{Cluster} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon,$$

where:

- Cluster is the outcome variable representing the resultant clusters of chronic conditions.

- β_0, β_1, β_2, ..., βn are the coefficients associated with the predictor variables X_1, X_2, ..., X_n.

- ε represents the error term.

This analysis will shed light on the factors contributing to the observed clusters and the extent of their impact, and whether these factors vary for each site.

Expected Results

The study seeks to provide essential insights and understandings into the clustering and prevalence of chronic conditions and their multimorbidity among different populations. We anticipate uncovering both common and site-specific clusters of multimobidites. This could help identify unique health challenges faced by different communities, shedding light on local and continental health disparities.

Furthermore, the analysis is expected to reveal how multimorbidity varies among different demographic groups. For example, older individuals may show higher instances of multiple chronic conditions, while distinctive patterns in younger demographics might highlight emergent public health issues. Such findings will enhance our understanding of the epidemiological transitions occurring in various regions and provide evidence to tailor public health policies and interventions effectively.

Behavioural and demographic factors like tobacco and alcohol use, marital status, and gender are also expected to play significant roles in the prevalence and types of multimorbidity observed. Understanding these relationships will be crucial for developing targeted prevention and management strategies, aiming to mitigate the impact of these risk factors on public health.

Lastly, this study will contribute to the advancement of statistical and analytical methodologies used in health research. Techniques such as hierarchical clustering, k-medoids clustering, and latent class analysis will be refined and validated through their application in this large-scale study, potentially setting standards for data analysis in health surveillance and epidemiology. The outcomes of this research will provide a framework for health practitioners and policymakers to implement more directed and effective health interventions, ultimately aiming to improve health outcomes and reduce the burden of chronic diseases across different populations in Africa.