# MULTIPLE-INDICATOR, MULTIPLE CAUSE MODELLING TO EXAMINE THE RELATIONSHIP BETWEEN FOODS CONSUMED AND NON-COMMUNICABLE DISEASES

**Laina Mbongo[1]*, Lawrence Kazembe, Lillian Pazvakawambwa**
Department of Computing, Mathematics and Statistical Sciences,
University of Namibia, Windhoek, Namibia

## Abstract

Non-Communicable diseases are commonly associated with the dietary patterns of an individual. quantifying the disease's burden over a household's or individual's health has been a topic of great interest to researchers as well as policymakers. Various measurement approaches of NCD's that account for different types of biasness is required to correctly identify explanatory variables. This chapter used Namibia Household and Income Expenditure (NHIES) survey of 2015/16 variables to examine relationships between NCDs and the type of foods consumed. Principal Component Analysis was used as a data reduction method to derive dietary patterns. Furthermore, this chapter applied a Multiple-Indicator, Multiple-Cause (MIMIC) model in which NCD's is dealt with as an unobserved construct or latent variable to be determined by its causes and indicators and to be estimated in a system of structural equations. SEM was used to assess the association between the prevalence of NCD's and food types consumed. Fruits, foods such as condiments/tea/coffee and potatoes, yams, cassava, or any foods made from roots and tubers accounted for majority of the variation. The SEM showed that food types such as local grains, meat and food made from oil or were found to be significant at 5% level.

**Keywords:** Structural Equation Models (SEM), Principal Component Analysis (PCA), Non-Communicable Diseases (NCD)

---

[1]* Corresponding Author Email: inambongo@gmail.com

## 1.1. Introduction

Non-communicable diseases (NCDs) kill about 41 million people each year, equivalent to 71% of all deaths globally. Each year, more than 15 million people die from a non-communicable disease between the ages of 30 and 69 years; 85% of these "premature" deaths occur in low- and middle-income countries (WHO, 2021). Non-communicable diseases, sometimes referred to chronic diseases, tend to be of long duration and are the result of a combination of genetic, physiological, environmental and behavioural factors. The main types of NCD are cardiovascular diseases (such as heart attacks and stroke), cancers, chronic respiratory diseases (such as chronic obstructive pulmonary disease and asthma) and diabetes (WHO, 2021).

The prevalence of multimorbidity is increasing worldwide. A systematic review in WHO Eastern Mediterranean countries in 2013 showed that the high mortality of NCDs is partially related to their multimorbidity. More than half of the adults with NCDs have multimorbidity or multiple concurrent morbid conditions, and not one single chronic disease (Khorrami, et al., 2020). The NCD's are mostly driven by forces that include rapid unplanned urbanization, globalization of unhealthy lifestyles and population ageing. Increased prevalence of obesity, increased consumption of poor-quality diets, and pervasive undernutrition are contributing to this epidemic (UNSCN, 2018).

Poor quality diets are found to be among the top 6 risk factors contributing to the global burden of disease (Global Pattern, 2016). According to Global Pattern (2016), the NCD burden is specifically associated with diets that are low in fruits and vegetables, high in sodium, low in nuts and seeds, low in whole grains, and low in seafood-derived omega-3 fatty acids. The type of dietary pattern followed can easily influence one's health and the risk of contracting a chronic illness. Three (3) categories of dietary pattern analysis approaches exist, namely the theoretical methods, empirical methods and the hybrid methods.

During the past few decades, quantifying the disease's burden over the population's health has been a topic of great interest to researchers as well as policymakers. A great deal of research has been conducted in the developed world to quantify the disease burden (communicable and non-communicable) on the population's health (El-Saadani, Saleh, & Ibrahim, 2021). One type of approach is based on Multiple-Indicator, Multiple-Cause (MIMIC) models in which NCD's is dealt with as an unobserved construct or latent variable to be determined by its causes and indicators and to be estimated in a system of structural equations. Measurement of Multiple-Indicator, Multiple-Causes variables such as prevalence of Non-Communicable Diseases or number of food groups consumed in a household can be a challenge to compute. In most instances, traditional analysis using a multivariate normal approximation for such type of variables can be misleading due to the nature of the data (small marginal means with a lot of zero counts) (Karlis & Meligkotsidou, 2007). Structural Equation Models combines both measurement and structural considerations. They integrate psychometric concepts (i.e., measurement approaches) and econometric ideas (structure approaches). Thus, this method has the ability to take into account measurement errors. As for the structure approaches in SEM, path analysis is applied to estimate the relationships among latent constructs. The ability to combine these two analyses is one of the advantages of SEM. By specifying and describing the plausible relationships between latent concepts and manifest variables, associated measurement errors, and proposed structural relationships among latent structures in SEM can effectively estimate parameters simultaneously, which mirror the fact that the variables coexist in reality.

Theoretical methods are also known as a priori methods and are used to assess diets based on prior knowledge and scientific evidence such as the dietary guideline index (Castro, Baltar, & Marchioni, 2016). Dietary indices are the most common hypothesis-oriented approaches that evaluate the adherence of population intake to nutritional recommendations. The common dietary indices include the Healthy Eating Index (HEI) that was developed to investigate American eating habits and their compliance with the dietary guidelines as provided by the Recommended Dietary Allowance (RDA) (de Calvalho, Dutra, Pizato, Gruezo, & Ito, 2014); the Original Diet Quality index that was developed to assess the intake of eight food groups and the recommendations of the committee on diet and health (Patterson et al., 1994); the Mediterranean diet score that is characterized by high intake of olive oil, non-starchy vegetables, legumes, whole grains, fruits and the low intake of whole milk and dairy products and red meats; and low to moderate intake of wine as the main source of alcohol during the meals (de Calvalho, Dutra, Pizato, Gruezo, & Ito, 2014); the Overall Nutritional Quality Index for assessing the overall nutritional quality of foods, and the Dietary Approaches to Stop Hypertension (DASH) which is a lifelong approach to healthy eating that is designed to help treat or prevent high blood pressure.

Empirical methods, sometimes referred to as a posteriori, uses statistical approaches to deduce information about existing dietary patterns within the population (Thorpe, Milte, Crawford, & McNaughton, 2016). Exploratory factor analysis is used to analyse interrelationships among a large number of variables and to explain these variables in terms of smaller number of common underlying dimensions. It involves finding a way of shrinking the information contained in some of the original variables into a smaller set of implicit variables with a minimal loss of information (Zaiontz, 2018). Principal Component Analysis (PCA) and Cluster Analysis (CA) are the other commonly used empirical methods for dietary patterns.  PCA uses the correlation matrix of food intake variables to identify common patterns of food consumption within the data to account for the largest amount of variation in diet (Thorpe, Milte, Crawford, & McNaughton, 2016). Both PCA and factor analysis are most suitable when confronted with a large number of correlated variables, and the desire is to reduce them into a small set of non-correlated variables that contains the same information of the larger one. Other reduction methods include the Cluster analysis, the Least Absolute Shrinkage and Selection Operator (LASSO), Reduced Rank Regression (RRR), and the partial least-squares regression.

This chapter thus aims to apply structural equation models to multiple-indicator, multiple causes dataset. The model is used to find the relationship between non-communicable diseases (NCD's) and the type of diets consumed in Namibia. Additionally, the chapter explored other data reduction method, PCA, to explain the type of foods consumed.

## 1.2. Materials and Methods

### 1.2.1. The NHIES 2015/16

The study used cross-sectional survey data of the Namibian Household and Income Expenditure (NHIES) of 2015/2016. The primary sampling frame that was used for this survey is a list of Primary sampling Units (PSUs) based on the 2011 Population and Housing Census Enumeration Areas (EAs). A secondary sampling frame for each of the selected PSUs was created for the purpose of selecting the sample households through a listing procedure. The sample design for the survey was a stratified two-stage cluster sample, where the first stage units were geographical areas designated as the Primary Sampling Units (PSUs) and the second stage units were the households. The up-to-date list of households in the selected PSU were prepared during the listing stage of fieldwork, and 12 households were systematically selected in each PSUs.

For this analysis, five (5) non-Communicable diseases; Diabetes (0.8%), High Blood Pressure (6.7%), Cancer (0.2%), Cardiac/Heart diseases (0.8%) and respiratory diseases (including asthma) (1.5%) were selected for analysis due to their high prevalence. Structural Equation Models (SEM) were used to model for NCD's, and the type of foods consumed. The food groups in the NHIES 2015/2016 were re-grouped and re-arranged in order to make up the 12 food groups. Principal Component analysis was used to reduce the 12 food groups to a few principal components. SPSS & statistical R Version 3.6 was used to compute PCA and SEM, respectively.

### 1.2.2. Statistical Methods

#### 1.2.2.1. Principal Component Analysis (PCA)

Two extensively used empirical methods for food pattern analysis are principal component analysis (PCA) and cluster analysis (CA) (Thorpe, Milte, Crawford, & McNaughton, 2016). In order to find common patterns of food consumption within the data and account for the most variation in diet, PCA uses the correlation matrix of food intake variables (Thorpe, Milte, Crawford, & McNaughton, 2016). PCA and factor analysis are mostly used when there are a large number of potential variables to analyze and there is a need to summarize the information contained in those variables as efficiently as possible (Gleason, Boushey, Harris, & Zoellner, 2015). The following definition for PCA is derived from (Zaiontz, 2018)

Let $X = [x_i]$ be any $k \times 1$ random vector. We now define a $k \times 1$ vector $Y = [y_i]$, where for each $i$ the $i_{th}$ principal component of $X$ is

$$y_i = \sum_{j=1}^{k} \beta_{ij} x_j \tag{1}$$

for some regression coefficients $\beta_{ij}$. Since each $y_i$ is a linear combination of the $x_j$, $Y$ is a random vector.

Let $\sum = [\sigma_{ij}]$ be the $k \times k$ population covariance matrix for $X$. Since the column vectors $\beta_j$ are orthonormal, $\beta_i . \beta_j = \beta_i^T \beta_j = 0$ if $j \neq i$ and $\beta_i^T \beta_j = 1$ if $j = i$. Then the covariance matrix for $Y$ is given by:

$$var(y_i) = \sum_{p=1}^{k} \sum_{m=1}^{k} \beta_{ip} \beta_{im} \sigma_{pm} = \beta_i^T \left( \sum_{j=1}^{k} \lambda_j \beta_j \beta_j^T \right) \beta_i = \sum_{j=1}^{k} \lambda_j \left( \beta_i^T \beta_j \right) \left( \beta_j^T \beta_i \right) = \lambda_i \quad (2)$$

$$cov\left(y_i, y_j\right) = \sum_{p=1}^{k} \sum_{m=1}^{k} \beta_{ip}\beta_{jm}\sigma_{pm} = \beta_i^T \left(\sum_{r=1}^{k} \lambda_r \beta_r \beta_r^T\right)\beta_j = \sum_{r=1}^{k} \lambda_r \left(\beta_i^T \beta_r\right)\left(\beta_r^T \beta_j\right) = 0 \quad (3)$$

It is also worth noting that the first principal component is the combination that accounts for the largest variance in the sample. The second component accounts for the next largest amount of variance and is uncorrelated with the first. Successive components thus explain progressively smaller portions of the sample variance and are uncorrelated with each other (Suresh., 2014).

Since one can calculate as many principal components as there are variables, the researcher does not gain any additional insight if all the variables are replaced by their principal components. Thus, one needs to determine how many factors are needed to represent the data, i.e., to reproduce the original correlations. There are two main criteria for deciding how many factors to extract. One by examining Eigenvalues whereby a criterion of eigenvalue greater than 1 suggests that only factors that account for variances greater than 1 should be included. Factors with a variance of less than 1 are not better than individual variables, since each variable has a variance of 1. Additionally, they can be studied using a scree plot, which plots the eigenvalues versus the number of variables in the order of extraction. The curve's point where the slope changes to a horizontal angle determines how many factors can be derived. The maximum number of components that can be extracted is indicated at this stage (Suresh ., 2014).

The other recommended method is the Varimax Orthogonal Rotation. The VARIMAX method of rotation is the most frequently used rotation method (Hair et al., 1998, as cited in (Suresh., 2014)). It minimizes the number of variables that have high loadings on a factor, so that the factors can be interpreted more easily. The relationship between the test points remains the same as before. However, the axes are altered to interpret the factors more easily (Suresh., 2014).

### 1.2.2.2. Structural Equation Model (SEM)

Structural Equation Modelling abbreviated as SEM, is a very general statistical modelling technique, which is widely used in the behavioural sciences (Hox, Moerbeek, & Van De Schoot, 2017). It can be viewed as a combination of factor analysis and regression or path analysis. The interest in SEM is often on theoretical constructs, which are represented by the latent factors. The relationship between the theoretical constructs is represented by regression or path coefficients between the factors. The structural equation model implies a structure for the covariances between the observed variables, which provides the alternative name covariance structure modelling. It should be noted that the model can be extended to include means of observed variables or factors in the model, which makes covariance structure modelling a less accurate name (Hox et al., 2017).

Bardenheier, et al., (2013) used structural equation modeling with factor analysis, which groups inter-correlated variables into a single factor or latent construct, and path analysis, which includes the direct and indirect effects of factors previously reported associated with prediabetes. Direct effects are depicted as an arrow emanating from an independent variable (exposure) leading and pointing to a dependent variable (outcome). An indirect effect is depicted as a mediating variable having an arrow pointing to it from an independent variable but also pointing to yet another dependent variable. A confounder is depicted as a variable with direct effects on both the exposure and the dependent variable. Correlations between the measurement errors of two variables are represented by two-headed curving arrows, in which case only the measurement error terms are correlated.

Latent variable models typically have several indicators for each latent construct, the capacity to test models with multiple dependent variables, and the advantage of testing multiple integrated models simultaneously rather than factors one at a time. Additionally, structural equation modeling studies the direct and indirect impacts of mediators on dependent variables as well as complex associations between various mediators (Bardenheier, et al., 2013). Equally, in a traditional regression model, mediators would not be included because they would block the pathway between the independent variable of interest and the dependent variable. Thus, in the structural-equation model, the independent factors and combined mediated relationships can be examined simultaneously, determining the impact of each of the dependent variables in the appropriate order. Thus, the SEM includes mediating effects without sacrificing indirect effects of interest. For each relationship in the SEM model, only data missing for either the independent or dependent variable would be missing from that equation (Bardenheier, et al., 2013).

The latent variable is divided into two parts namely the latent variable model and the measurement model. The latent variable is defined as follows:

$$\eta_i = \alpha_\eta + \boldsymbol{B}\eta_i + \boldsymbol{\Gamma}\xi_i + \varsigma_i \qquad (4)$$

Whereby $\eta_i$ is a vector of latent endogenous variables for unit $i$, $\alpha_\eta$ is a vector of intercept terms for equations, $\boldsymbol{B}$ is the matrix of coefficients giving the expected effects of the latent endogenous variables ($\eta$) on each other, $\xi_i$ is the vector of latent exogenous variables, $\boldsymbol{\Gamma}$ is the coefficient matrix giving the expected effects of the latent exogenous variables ($\varsigma$) on the latent endogenous variables ($\eta$), and $\varsigma_i$ is the vector of disturbances. The $i$ subscript indexes the ith case in the sample.

The measurement model links the latent to the observed responses (indicators). It has two equations as outlined below:

$$y_i = \alpha_y + \Lambda_y\eta_i + \varepsilon_i \text{ and} \qquad (5)$$

$$x_i = \alpha_x + \Lambda_x\xi_i + \delta_i \qquad (6)$$

Where $y_i$ and $\mathrm{x_i}$ are vectors of the observed indicators of $\eta_i$ and $\xi_i$, respectively, $\alpha_y$ and $\alpha_{\mathrm{x}}$ are intercept vectors, $\Lambda_y$ and $\Lambda_x$ are matrices of factor loadings or regression coefficients giving the impact of the latent $\eta_i$and $\xi_i$ on $y_i$ and $\mathrm{x_i}$, respectively, and $\varepsilon_i$ and $\delta_i$ are the unique factors of $y_i$ and $\mathrm{x_i}$.

#### 1.2.2.3. Model Selection

Chi-square test statistics is the most used when modelling latent variables to measure/quantify model fit; however, it is sensitive to large sample size. Methodologists developed numerous fit indices to adjust the chi-square test statistics with the information in the model, such as degrees of freedom, sample size, and/or the number of variables. Chi-square can be calculated as follows:

$$X^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

Depending on the elements in the formula, fit indices in latent variable models can be categorized into three types (Chang, Gardiner, Houang, & Yu, 2020): 1) relative fit indices, Comparative Fit Index and the absolute fit indices). R packages: “lavaan” and “semPlot” were used to model the structural equations.

## 1.3. Results

### 1.3.1. Prevalence of Non- Communicable Diseases

Non-communicable diseases are a concern in Namibia. High Blood pressure was found to be highest (6.7%) NCD among the population in Namibia (Table 34). Other NC diseases that most people are suffering from diseases include Asthma and epilepsy (1.0%), diabetes and cardiac/heart diseases (0.8%) respectively, and cancer (0.2%).

**Table 1:** non-communicable diseases in Namibia

| Disease | Frequency | % |
|---|---|---|

| | | |
|---|---:|---:|
| **NCD1: Diabetes** | 348 | 0.8 |
| **NCD2: High blood pressure** | 2785 | 6.7 |
| **NCD3: Cancer** | 72 | 0.2 |
| **NCD4: Cardiac / Heart** | 336 | 0.8 |
| **NCD5: Respiratory disease (asthma, etc.)** | 641 | 1.5 |
| **NCD6: No Chronic illness** | 37399 | 89.9 |
| **Total** | **41581** | **100** |

### 1.3.2. Types of Food Consumed

Increased prevalence of obesity, increased consumption of poor-quality diets, and pervasive undernutrition are contributing to the NCD epidemic. Specifically, the NCD burden is associated with diets low in fruits and vegetables, high in sodium, low in nuts and seeds, low in whole grains, and low in seafood-derived omega-3 fatty acids (UNSCN, 2018). Table 35 shows that 23.4% of the food consumed are local foods mostly made from wheat or grain, 19.6% of the food consumed were from foods made with oil, fat or butter, 18.3% is meat products and 17.9% from sugar or honey. High consumption of these foods is associated with NCD’s (UNSCN, 2018).

**Table 2:** Type of foods consumed.

| Food Type | Yes | | No | |
|---|---:|---:|---:|---:|
| | Frequency | % | Frequency | % |
| Any (local food) bread, rice, noodles, biscuits or any other foods made from millet, sorghum, maize, rice, wheat, or (any other local grain) | 9739 | 23.40% | 351 | 0.80% |
| Beef, pork, lamb, goat, rabbit, wild game, chicken, duck, other birds, liver, kidney, heart or other organ meats | 7663 | 18.30% | 2427 | 5.80% |
| Foods made with oil, fat or butter | 8157 | 19.60% | 1933 | 4.60% |
| Sugar/honey | 7428 | 17.90% | 2662 | 6.40% |

### 1.3.3. Association of Type of Foods Consumed and Non-Communicable Diseases

#### 1.3.3.1. Local grain foods

At least 6.9% of the households with High Blood pressure have indicated that they consume “Any (local food) bread, rice, noodles, biscuits or any other foods made from millet, sorghum, maize, rice, wheat, or (any other local grain)”. Other NCD with A high percentage in consumption of local foods/grain/wheat was the respiratory diseases (including asthma) (1.4%). Additionally, the Pearson’s Chi-square test indicated that there was no association between NCDs and the food type (“Any (local food) bread, rice, noodles, biscuits or any other foods made from millet, sorghum, maize, rice, wheat, or (any other local grain)”), P-value 0.780. (Table 36).

**Table 3:** Association of NCD and Local Food

| **Disease** | **Any local food/grain/wheat** | | **Total** | **Pearson Chi-Square** | |
|---|---|---|---|---|---|
| | **No** | **Yes** | | **Value** | **Asymptotic Significance (2-sided)** |
| Diabetes | 0.90% | 0.80% | 0.80% | 3.228 | 0.78 |
| High Blood Pressure | 6.00% | 6.90% | 6.90% | | |
| Cancer | 0.00% | 0.20% | 0.20% | | |
| Cardiac or Heart | 1.40% | 0.90% | 0.90% | | |
| Respiratory Disease (Inc. Asthma) | 2.00% | 1.40% | 1.50% | | |
| Does not have a Chronic illness | 87.70% | 87.40% | 87.40% | | |
| **Total** | **100.00%** | **100.00%** | **100.00%** | | |

### 1.3.3.2. Meat Products

At least 7.1 percent of individuals with high blood pressure consumed meat/chicken products. Meat products included beef, pork, lamb, goat, rabbit, wild game, chicken, duck, other birds, liver, kidney, heart or other organ meats. The Pearson's Chi-square test shows that there is a significant relationship between the NCDs and Meat/Chicken products (P-value- 0.0034).

**Table 4:** Association of NCD and Meat

| **Disease** | **Meat/Chicken Products** | | **Total** | **Pearson Chi-Square** | |
|---|---|---|---|---|---|
| | **No** | **Yes** | | **Value** | **Asymptotic Significance (2-sided)** |
| Diabetes | 0.90% | 0.70% | 0.80% | 10.837 | 0.034 |
| High blood pressure | 6.20% | 7.10% | 6.90% | | |
| Cancer | 0.10% | 0.20% | 0.20% | | |
| Cardiac/Heart | 0.90% | 0.90% | 0.90% | | |
| Respiratory disease (asthma, etc.) | 1.00% | 1.60% | 1.50% | | |
| Does not have a Chronic illness | 88.70% | 87.00% | 87.40% | | |
| **Total** | **100.00%** | **100.00%** | **100.00%** | | |

### 1.3.3.3. Foods made with Oil, Fat or Butter

Table 38 indicates that 6.9%, 1.4%, 0.8%, 0.8% and 0.2% of the households had High Blood pressure, Respiratory diseases, Cardiac/ Heart, Diabetes, and cancer respectively (Table 38). The Pearson Chi-square was however not significant at 5% and indicated no association between the NCDs and foods made with oil, fat or butter.

**Table 5:** Association of NCD with Fats/Oils

| **Disease** | **Foods made with oil, fat or butter** | | **Total** | **Pearson Chi-Square** | |
|---|---|---|---|---|---|
| | **No** | **Yes** | | **Value** | **Asymptotic Sign. (2-sided)** |
| Diabetes | 0.80% | 0.80% | 0.80% | 8.537 | 0.201 |
| High blood pressure | 6.80% | 6.90% | 6.90% | | |
| Cancer | 0.30% | 0.20% | 0.20% | | |
| Cardiac/Heart | 1.40% | 0.80% | 0.90% | | |
| Respiratory disease (asthma, etc.) | 1.60% | 1.40% | 1.50% | | |
| Does not have a chronic illness | 86.50% | 87.70% | 87.40% | | |
| **Total** | 100.00% | 100.00% | 100.00% | | |

### 1.3.3.4. Sugar/Honey

Table 39 shows that at least 7.0 percent of the high blood pressure, 1.5% respiratory diseases, 0.9% Cardiac/Heart, 0.8% Diabetes and 0.2% Cancer individuals consumed sugar or honey products. According to the Pearson Chi-Square test, there was no association between NCDs and Sugar/Honey (P-value greater than 0.005).

**Table 6:** Association of NCD and Sugar/Honey

| **Disease** | **Sugar/Honey** | | **Total** | **Pearson Chi-Square** | |
|---|---|---|---|---|---|
| | **No** | **Yes** | | **Value** | **Asymptotic Sign. (2-sided)** |
| Diabetes | 0.60% | 0.80% | 0.80% | 12.273 | 0.056 |
| High blood pressure | 6.50% | 7.00% | 6.90% | | |
| Cancer | 0.10% | 0.20% | 0.20% | | |
| Cardiac/Heart | 1.00% | 0.90% | 0.90% | | |
| Respiratory disease (asthma, etc.) | 1.30% | 1.50% | 1.50% | | |
| Does not have a chronic illness | 87.30% | 87.50% | 87.40% | | |
| **Total** | **100.00%** | **100.00%** | **100.00%** | | |

### 1.3.4. Principal Component Analysis (PCA)

Principal Component analysis was used to reduce the 12 food groups to a few principal components. The PCA extracted three (3) components with eigen values greater than 1, explaining 49.4% of the total variance in the data set. The first, second and third components explained 29.7%, 10.4% and 9.4% respectively of all variations (Table 40).

**Table 7:** PCA components

| **Component** | **Initial Eigenvalues** | | | **Extraction Sums of Squared Loadings** | | **Rotation Sums of Squared Loadings**$^a$ | |
|---|---|---|---|---|---|---|---|
| | **Total** | **% of Variance** | **Cumulative %** | **Total** | **% of Variance** | **Cumulative %** | **Total** |
| 1 | 3.560 | 29.670 | 29.670 | 3.560 | 29.670 | 29.670 | 3.085 |
| 2 | 1.246 | 10.383 | 40.053 | 1.246 | 10.383 | 40.053 | 2.038 |
| 3 | 1.125 | 9.374 | 49.426 | 1.125 | 9.374 | 49.426 | 1.500 |
| 4 | .902 | 7.516 | 56.943 | | | | |
| 5 | .852 | 7.103 | 64.046 | | | | |
| 6 | .750 | 6.247 | 70.294 | | | | |
| 7 | .735 | 6.126 | 76.419 | | | | |
| 8 | .629 | 5.243 | 81.662 | | | | |
| 9 | .581 | 4.842 | 86.504 | | | | |
| 10 | .555 | 4.624 | 91.128 | | | | |
| 11 | .536 | 4.470 | 95.598 | | | | |
| 12 | .528 | 4.402 | 100.000 | | | | |
| Extraction Method: Principal Component Analysis. | | | | | | | |
| a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance | | | | | | | |

Component 1, 2 and 3 accounted for majority of the variances and had eigen values of 3.6, 1.2 and 1.1 respectively. This implies the PCA explained 49.9% of the food types summarized as three (3) underlying dimensions coined from the food types loaded significantly in the 3 extracted Components. Table 41 indicates that the variables can be grouped into three (3) components with three (3) factor loadings each. Food made with oil, fat and butter, and vegetables food items overlaps across the components but with their strongest loading of 0.516 and 0.477 respectively in the 1st component.

**Table 8:** Component Matrix of the PCA

| **Food Types** | **Component** | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| Fruits | .682 | -.334 | |
| Any foods such as condiments/tea/coffee | .672 | | |
| Potatoes, yams, cassava, or any foods made from roots and tubers | .670 | | |
| Eggs | .657 | | |
| Cheese, yoghurt, milk or other milk products | .605 | | -.322 |
| Sugar/honey | .597 | .450 | |
| Beef, pork, lamb, goat, rabbit, wild game, chicken, duck, other birds, liver, kidney, heart or other organ meats | .552 | | |
| Foods made with oil, fat or butter | .516 | .439 | .308 |
| Vegetables | .477 | -.320 | .380 |
| Food made from beans, peas, lentils or nuts | .319 | -.448 | |
| Any (local food) bread, rice, noodles, biscuits or any other foods made from millet, sorghum, maize, rice, wheat, or (any other local grain) | | .413 | |
| Fresh or dried fish or shellfish | | | .733 |
| Extraction Method: Principal Component Analysis. | | | |
| a. 3 components extracted. | | | |

Figure 3 shows graphic dimensions to determine the number of components to be extracted. The Scree plot suggests taking the first 3 components.

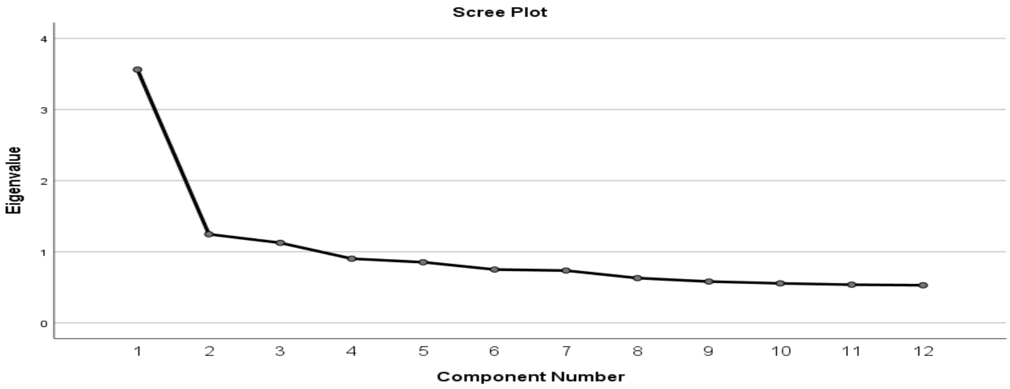

**Figure 1:** Scree Plot of food types

### 1.3.5. Structural Equation Modelling (SEM)

Our model is estimated by maximum likelihood (ML). A likelihood ratio statistic comparing the fitted model (with 75 parameters) to the unconstrained saturated model produces a p value of less than 0.001 with df = 166. The chi square test is significant suggesting good fit. Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI) are used to assess model fit. In this study, SRMR = 0.063, CFI= 0.850 and TLI= 0.799, which are near the recommended cutoffs of less than 0.08 and more than 0.95, indicating that the model is a good-fitting model. Comparative indices, such as AIC and BIC, are used to compare competing models, that is, with different co-variance structures. Since there is no competitive model in our example, AIC and BIC is not used in this case.

**Table 9:** SEM Model Specifications

| **Estimator** | **ML** |
|---|---|
| **Optimization method** | NLMINB |
| ***Model Test User Model*** | |
| **Test Statistic** | 6211.565 |
| **D.F** | 168 |
| **P-Value** | <0.001 |
| ***Model Test Baseline Model*** | |
| **Test Statistic** | 40403.294 |
| **D.F** | 225 |
| **P-Value** | <0.001 |

| | |
|---|---|
| ***User Model versus Baseline Model*** | |
| **Comparative Fit Index (CFI)** | 0.850 |
| **Tucker-Lewis Index (TLI)** | 0.799 |
| ***Log likelihood and Information criterion*** | |
| **Log likelihood User Model (H0)** | -9079.044 |
| **Log likelihood unrestricted model (H1)** | -5973.261 |
| **Akaike (AIC)** | 18308.087 |
| **Bayesian (BIC)** | 18849.535 |
| **Sample-size adjusted Bayesian** | 18611.196 |
| ***Root Mean Square Error of Approximation*** | |
| **RMSEA** | 0.060 |
| **90% CU – Lower** | 0.058 |
| **90% CI- Upper** | 0.061 |
| **P-value RMSEA <=0.05** | 0.000 |
| ***Standardized Root Mean Square Residual*** | |
| ***SRMR*** | 0.063 |
| ***Parameter Estimates*** | |
| **Standard errors** | Standard |
| **Information** | Expected |
| **Information saturated (h1) model** | Structured |

R packages 'lavaan" and "semPlot" used to run structural equation modeling in this chapter (Table 43, 44, 45, 46), and the latter one is for generating the diagram (Figure 4). The "Std.lv" column reported the estimates when the latent variables "FS" (Food Security) and "NCD" (Non-Communicable Diseases) were standardized. The last column "Std.all" reported the parameter estimates when both the latent variables and the observed variables were standardized (also called the 'completely standardized solution'). The function semPaths is used to plot the SEM diagrams (Figure 4).

**Table 10:** Parameter Estimates: Latent Variables

| **Latent variables. FS=~** | **Estimate** | **Std. err** | **P value** | **Std. lv** | **Std. all** |
|---|---|---|---|---|---|
| **Food Type 1** | 1.000 | | | 0.038 | 0.206 |
| **Food Type 2** | 7.975 | 0.435 | <0.001 | 0.301 | 0.629 |
| **Food Type 3** | 5.303 | 0.310 | <0.001 | 0.200 | 0.405 |
| **Food Type 4** | 7.829 | 0.427 | <0.001 | 0.295 | 0.636 |
| **Food Type 5** | 5.443 | 0.308 | <0.001 | 0.205 | 0.481 |
| **Food Type 6** | 7.179 | 0.393 | <0.001 | 0.271 | 0.613 |
| **Food Type 7** | 2.718 | 0.203 | <0.001 | 0.103 | 0.205 |
| **Food Type 8** | 2.537 | 0.168 | <0.001 | 0.096 | 0.265 |
| **Food Type 9** | 7.263 | 0.402 | <0.001 | 0.274 | 0.553 |
| **Food Type 10** | 4.450 | 0.257 | <0.001 | 0.168 | 0.427 |
| **Food Type 11** | 6.071 | 0.339 | <0.001 | 0.229 | 0.520 |
| **Food Type 12** | 0.154 | 0.446 | <0.001 | 0.308 | |

| | | | | | |
|---|---|---|---|---|---|
| **Ncd1=~ncd_1 (Diabetes)** | 1.000 | | | 0.087 | 1.000 |
| **Ncd2=~ncd_2 (High blood pressure)** | 1.000 | | | 0.253 | 1.000 |
| **Ncd3=~ncd_3 (Cancer)** | 1.000 | | | 0.044 | 1.000 |
| **Ncd4=~ncd_4 (Cardiac/Heart)** | 1.000 | | | 0.095 | 1.000 |
| **Ncd5=~ncd_5 (Respiratory illness)** | 1.000 | | | 0.120 | 1.000 |
| **Ncd6=~ncd_6 (No NCD)** | 1.000 | | | 0.331 | 1.000 |

Table 44 shows regression estimates. Educational level of a household and residence type was found to have a statistically significant relationship (p-value less than 0.001) with diabetes (NCD1). Other variables that were significant at 5% are residence (Diabetes (NCD1), Cardiac/Heart (NCD4) and No NCD (NCD6)). Table 45 show covariances of NCD’s.

**Table 11:** Parameter Estimates: Regression

| **Variable** | **Estimate** | **Std. err** | **P value** | **Std. lv** | **Std. all** |
|---|---|---|---|---|---|
| ***Ncd1 (Diabetes)~*** | | | | | |
| **Food Security** | -0.042 | 0.026 | 0.100 | -0.018 | -0.018 |
| **Attain (educational level)** | 0.005 | 0.001 | <0.001 | 0.060 | 0.060 |
| **li_urbrur (residence)** | -0.000 | 0.000 | 0.011 | -0.001 | -0.026 |
| **q04_20 (Smoking)** | -0.005 | 0.003 | 0.105 | -0.056 | -0.017 |
| **Q04_22 (Alcohol consumption)** | 0.002 | 0.002 | 0.315 | 0.025 | 0.011 |
| ***Ncd2 (High Blood Pressure) ~*** | | | | | |
| **Food Security** | 0.022 | 0.073 | 0.760 | 0.003 | 0.003 |
| **Attain (educational level)** | 0.035 | 0.003 | <0.001 | 0.140 | 0.139 |
| **li_urbrur (residence)** | 0.000 | 0.000 | 0.167 | 0.000 | 0.014 |
| **q04_20 (Smoking)** | -0.040 | 0.009 | <0.001 | -0.157 | -0.048 |
| **Q04_22 (Alcohol consumption)** | 0.007 | 0.0060 | 0.260 | 0.028 | 0.012 |
| ***Ncd3 (Cancer) ~*** | | | | | |
| **Food Security** | 0.021 | 0.013 | 0.109 | 0.018 | 0.018 |
| **Attain (educational level)** | 0.002 | 0.000 | <0.001 | 0.038 | 0.038 |
| **li_urbrur (residence)** | -0.000 | 0.000 | 0.796 | -0.000 | -0.003 |
| **q04_20 (Smoking)** | -0.003 | 0.002 | 0.045 | -0.069 | -0.021 |
| **Q04_22 (Alcohol consumption)** | 0.003 | 0.001 | 0.045 | 0.058 | 0.025 |
| ***Ncd4 (Cardiac/heart) ~*** | | | | | |
| **Food Security** | 0.013 | 0.028 | 0.635 | 0.005 | 0.005 |
| **Attain (educational level)** | 0.002 | 0.001 | 0.085 | 0.018 | 0.017 |
| **li_urbrur (residence)** | 0.000 | 0.000 | 0.003 | 0.001 | 0.030 |
| **q04_20 (Smoking)** | -0.001 | 0.003 | -0.373 | -0.013 | -0.004 |
| **Q04_22 (Alcohol consumption)** | -0.000 | 0.002 | -0.083 | -0.002 | -0.001 |
| ***Ncd5 (Respiratory illness) ~*** | | | | | |
| **Food Security** | 0.044 | 0.035 | 0.208 | 0.014 | 0.014 |
| **Attain (educational level)** | 0.002 | 0.001 | 0.144 | 0.015 | 0.015 |
| **li_urbrur (residence)** | 0.000 | 0.000 | 0.248 | 0.000 | 0.012 |
| **q04_20 (Smoking)** | -0.008 | 0.004 | 0.046 | -0.069 | -0.021 |
| **Q04_22 (Alcohol consumption)** | 0.004 | 0.003 | 0.224 | 0.030 | 0.012 |
| ***Ncd6 (No NCD) ~*** | | | | | |
| **Food Security** | 0.009 | 0.096 | 0.092 | 0.001 | 0.001 |
| **Attain (educational level)** | -0.047 | 0.003 | <0.001 | -0.142 | -0.141 |
| **li_urbrur (residence)** | -0.000 | 0.000 | 0.021 | -0.000 | -0.023 |

| | | | | | |
|---|---|---|---|---|---|
| **q04_20 (Smoking)** | 0.063 | 0.011 | <0.001 | 0.190 | 0.058 |
| **Q04_22 (Alcohol consumption)** | -0.017 | 0.008 | 0.037 | -0.051 | -0.022 |

**Table 12:** Parameter Estimates: Covariances

| Variable (NCD) | Estimate | Std.err | P value | Std.lv | Std.all |
|---|---|---|---|---|---|
| ***Ncd1~~*** | | | | | |
| **Ncd2** | -0.001 | 0.000 | **0.001** | -0.033 | -0.033 |
| **Ncd3** | -0.000 | 0.000 | 0.505 | -0.007 | -0.007 |
| **Ncd4** | -0.000 | 0.000 | 0.384 | -0.009 | -0.009 |
| **Ncd5** | -0.000 | 0.000 | 0.245 | -0.012 | -0.012 |
| **Ncd6** | 0.006 | 0.000 | **0.001** | -0.225 | -0.225 |
| ***Ncd2~~*** | | | | | |
| **Ncd3** | -0.000 | 0.000 | 0.064 | -0.018 | -0.018 |
| **Ncd4** | -0.001 | 0.000 | 0.004 | -0.029 | -0.029 |
| **Ncd5** | -0.001 | 0.000 | **<0.001** | -0.036 | -0.036 |
| **Ncd6** | -0.058 | 0.001 | **<.0001** | -0.710 | -0.710 |
| ***Ncd3~~*** | | | | | |
| **Ncd4** | -0.000 | 0.000 | 0.642 | -0.005 | -0.005 |
| **Ncd5** | -0.000 | 0.000 | 0.506 | -0.007 | -0.007 |
| **Ncd6** | -0.002 | 0.000 | <0.001 | -0.113 | -0.113 |
| ***Ncd4~~*** | | | | | |
| **Ncd5** | -0.000 | 0.000 | 0.225 | -0.012 | -0.012 |
| **Ncd6** | -0.008 | 0.000 | **<0.001** | -0.252 | -0.252 |
| ***Ncd5~~*** | | | | | |
| **NCd6** | -0.013 | 0.000 | **<0.001** | -0.321 | -0.321 |

Table 46 shows variances of food types and NCD's. all the food types were significant at 5%.

**Table 13:** Parameter Estimates: Variances

| Variable | Estimate | Std. err | P value | Std. lv | Std. all |
|---|---|---|---|---|---|
| **Food Type 1** | 0.032 | 0.000 | <0.001 | 0.032 | 0.958 |
| **Food Type 2** | 0.138 | 0.002 | <0.001 | 0.138 | 0.604 |
| **Food Type 3** | 0.204 | 0.003 | <0.001 | 0.204 | 0.836 |
| **Food Type 4** | 0.129 | 0.002 | <0.001 | 0.129 | 0.596 |
| **Food Type 5** | 0.140 | 0.002 | <0.001 | 0.140 | 0.769 |
| **Food Type 6** | 0.122 | 0.002 | <0.001 | 0.122 | 0.624 |
| **Food Type 7** | 0.239 | 0.003 | <0.001 | 0.239 | 0.958 |
| **Food Type 8** | 0.121 | 0.002 | <0.001 | 0.121 | 0.930 |
| **Food Type 9** | 0.170 | 0.003 | <0.001 | 0.170 | 0.694 |
| **Food Type 10** | 0.127 | 0.002 | <0.001 | 0.127 | 0.818 |
| **Food Type 11** | 0.142 | 0.002 | <0.001 | 0.142 | 0.730 |
| **Food Type 12** | 0.155 | 0.002 | <0.001 | 0.155 | 0.621 |
| **Ncd1=~ncd_1 (diabetes)** | 0.000 | | | 0.000 | 0.000 |
| **Ncd2=~ncd_2 (High blood pressure)** | 0.000 | | | 0.000 | 0.000 |
| **Ncd3=~ncd_3 (Cancer)** | 0.000 | | | 0.000 | 0.000 |
| **Ncd4=~ncd_4 (Cardiac/Heart)** | 0.000 | | | 0.000 | 0.000 |
| **Ncd5=~ncd_5 (Respiratory illness)** | 0.000 | | | 0.000 | 0.000 |
| **Ncd6=~ncd_6 (No NCD)** | 0.000 | | | 0.000 | 0.000 |
| **FS (Food Security)** | 0.001 | 0.000 | <0.001 | 1.000 | 1.000 |
| **NCDs** | 0.008 | 0.000 | <0.001 | 0.995 | 0.995 |
| **NCD2** | 0.063 | 0.001 | <0.001 | 0.979 | 0.979 |

| | | | | | |
|---|---|---|---|---|---|
| **NCD3** | 0.002 | 0.000 | <0.001 | 0.998 | 0.998 |
| **NCD4** | 0.009 | 0.000 | <0.001 | 0.999 | 0.999 |
| **NCD5** | 0.014 | 0.000 | <0.001 | 0.999 | 0.999 |
| **NCD6** | 0.107 | 0.002 | <0.001 | 0.978 | 0.978 |

Figure 4 shows structural equation modelling pathways between type of foods consumed, non-communicable diseases and other socio-economic variables.

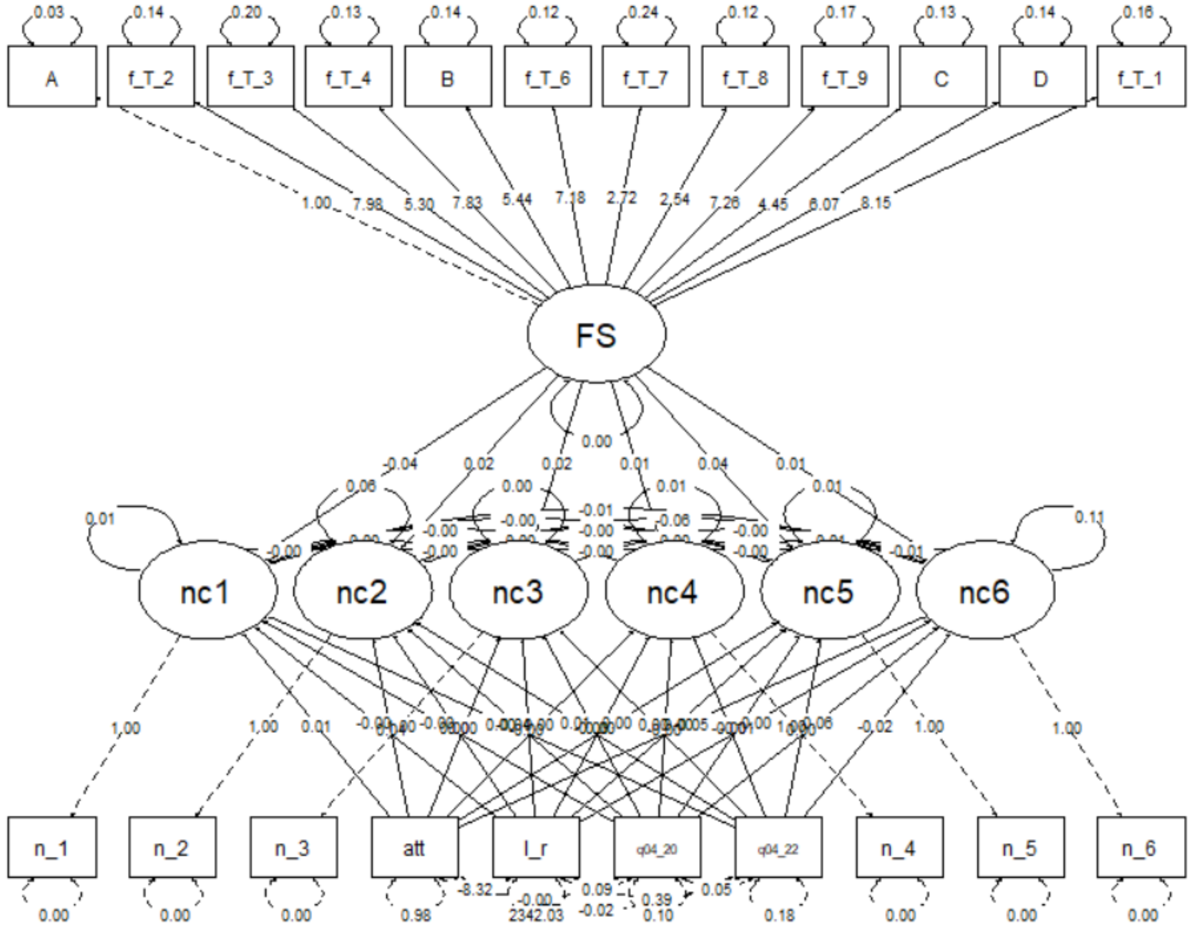

**Figure 2:** SEM: Foods Consumed, NCDs and Socio-economic variable

## 1.4. Discussion

This chapter modelled multiple indicator-multiple causes regression using Structured Equation Models (SEM). The study first looked at the prevalence of NCD's in the country. Non-communicable diseases have been on a rise, especially in low-middle income countries for the past decades (UNSCN, 2018). High blood pressure was particularly found to contribute largely to NCDs in the country. High Blood Pressure or sometimes referred to as hypertension, is a common disorder that affects a large proportion of the community. It is mostly asymptomatic and is detected on routine exams or after the occurrence of a complication such as a heart attack or stroke (Sunil & Gregory, 2021). Globally, the overall prevalence of hypertension in adults is estimated to be between 30-45%, with a higher prevalence in men than women (24% and 20%) respectively (Williams, Mancia, & Spiering, 2018). Other types of NCDs are respiratory illnesses, diabetes and cardiac or heart disease and cancer.

The types of food consumed has a significant contribution to the presence of an NCD in an individual. A diet that lacks fruits and vegetables and has a high intake of sodium, low intake of nuts and seeds, low in whole grains as well as in seafood-derived omega-3 fatty acids is specifically associated with a high prevalence of NCDs (Global Pattern, 2016). This chapter analysed the association between NCDs and foods types such as local food (bread, rice, noodles, biscuits or any other food made from millet, sorghum, maize, rice, wheat, or any other local grain), Meat (beef, pork, lamb, goat, rabbit, wild game, chicken, duck, other birds, liver, kidney, heart or other organ meats), Foods made with oil, fat or butter. Among all the food types, the meat was found to have a significant effect to NCDs. An analysis done by the American Heart Association found that each serving per day of processed meat was associated with a 42% higher risk of coronary heart disease and a 19% higher risk of diabetes, while total meat intake was associated with a 25% higher risk of coronary heart disease (Micha, Wallace, & Mozaffarian, 2010).

Empirical methods, particularly the Principal Component Analysis (PCA) as a data reduction method was used to derive dietary patterns. PCA used the correlation matrix of food intake groups to identify common patterns of food consumption in the dataset to account for the largest amount of variation in diet. Fruits, foods such as condiments/tea/coffee and potatoes, yams, cassava, or any foods made from roots and tubers accounted for majority of the variation. Furthermore, the SEM was derived through the packages "lavaan" and "semPlot" to analyse multivariate data with multiple indicators. Food types such as local grains, meat and food made from oil were found to be significant at 5% level and associated with NCDs. This concurs with the findings by Micha, Wallace, & Mozaffarian (2010).

## 1.5. Conclusion

Multiple data analysis has been on an increase for researchers for the past decades, but models accounting for multiple-indicators, multiple cause data are rarely applied for its computational difficulties. Mostly, for single counts, the Poisson regression model is used. The limitations of Poisson regression model are the assumption of equal mean and variance and restricting the count variables to positive. The aim of this study was to model multiple-indicator, multiple-cause to examine the relationship between foods consumed and NCDs. The study concluded that the type of food consumed has a significant contribution to the incidence of an NCD in an individual. In this chapter, we employed Structural Equation Models. The SEM analyses structural relationships between Non-Communicable Diseases and Foods Consumed and other Socio-economic variables.

## 1.6. Acknowledgements

The authors 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.