University of Fort Hare
Together in Excellence

Factor Analysis for Determination of Metabolic Syndrome Components on Anthropometric Data from Kinshasa Hinterland of the Democratic Republic of Congo

By

John S. Nasila

(201214428)

A Dissertation Submitted to the University of Fort Hare,
Faculty of Science and Agriculture, Department of Statistics
For the fulfilment of the requirements of the degree of Doctor of Philosophy in Biostatistics

Supervisor: Prof Y. Qin

Co – Supervisor: Dr J. Ndege

2017

Declaration

I, **John S Nasila** hereby declare that the content of this research work is my own and information extracted from other sources have been duly acknowledged as such. I further confirm that the research work has not been submitted in part or in full for an award of any other degree or in any other institution of higher learning apart from University of Fort Hare, South Africa.

Signature: 

JOHN S NASILA

Certification

This is to certify that this research work was carried out by **John S. Nasila**.
and submitted to the University of Fort Hare, Faculty of Science and Agriculture,
Department of Statistics for the fulfilment of the requirements of the degree of

Doctor of

Philosophy in Biostatistics

Signature:…… 21/08/ 2017

(John S. Nasila) Date

SUPERVISOR: Prof Y. Qin 21/08/ 2017

Signature Date

CO - SUPERVISOR: Dr J Ndege 21/08/ 2017

Signature Date

2017

Abstract

Factor analysis is a multivariate statistical approach commonly used in psychology, education, and more recently in the health-related professions. This thesis will attempt to provide novice and experienced researchers an application of two factor analysis methods which are; exploratory factor analysis (EFA) and Confirmatory Factor Analysis (CFA) to medical data. As Biostatistics knowledge continues to grow, it is timely that this thesis contributes immensely; firstly to the discipline of Biostatistics and secondly to Medicine both nationally and internationally. Factor analysis is an important tool that can be used in the development, refinement, and evaluation of tests, scales, and measures that can be used in education and in health-related professions such as medicine. This thesis is focused on applying Factor Analysis on medical data, specifically on data obtained from patients that suffer from Metabolic Syndrome and patients who don't suffer from Metabolic Syndrome.

Metabolic Syndrome (MS) is a constellation of components (factors) such as obesity, lipid-lipoprotein (fats) disorders, increase in glucose (sugar), hypertension (blood pressure), and inflammation/hypercoagulability (clotting). MS and other risk factors; (smoking, physical inactivity, excessive alcohol intake, and inappropriate diet) determine high morbidity and mortality for the cardiovascular disease (CVD=heart attack, brain attack, peripheral vascular disease) or cardio-metabolic risk (CMR=type 2 diabetes, kidney disease, retinopathy). Obesity, CVD, and CMR are emerging as epidemic conditions worldwide. However, Africa is not paying priority to early detection, treatment, prevention and control of atherosclerotic diseases (MS, CMR) from valid and reliable data.

The aim of this thesis was to examine anthropometry, glucose and blood pressure (non-lipid components of MS) as most valid, reliable, less time-consuming, less

complex and less expensive procedure of identifying people at high risks of CVD and CMR.

A further contribution of this thesis was its understanding of the economic implications of the burden of Metabolic Syndrome. Other burden factors have been identified and also discussed. The study has revealed that the presence of metabolic syndrome has contributed to an enormous economic burden by about 20% of the total economic loss experienced by many countries. The prevalence has risen recently and elevated patients' use of more health care resources, and face higher morbidity and mortality, resulting in an enormous economic burden. Some studies have shown healthcare costs to be as much as 20% higher than those accrued by patients without the risk factors. Patients with the Metabolic Syndrome have been shown to have greater drug expenditures, more frequent hospitalizations, and higher utilization of outpatient and physician services. When considered alone, the individual risk factor components account for a substantial economic burden to patients, health plans, and society as a whole. Overall, this has had serious economic impacts on many countries. The diagnosis of Metabolic Syndrome as a condition may encourage appropriate management and thus help prevent disease progression and reduce the considerable economic impact.

This study was a cross-sectional, comparative, and correlational survey conducted between January and April 2005, in Kinshasa Hinterland, DRC. Participants were black Bantu Africans.

In this study, the researcher attempted to determine latent factors that could explain the variability in a large set of data collected on many individuals of mixed health statuses. The original population consisted of 9770 people of whom, only 977 (10%) participated. Factor analysis and interpretation of the results were based on anthropometric parameters (body mass index or BMI and waist circumference or

WC), blood pressure (BP), lipid (triglycerides)-lipoprotein (HDL-C) and glucose with different numbers and cutoffs of components of Metabolic Syndrome.

A number of different statistical procedural methods have been employed to clearly scrutinize and bring out the information which is concealed in a variety of variables observed/collected on many human participants. A large portion of these approaches was based on multivariate statistical methods.

The approach, in this case, was the application of Principal Component Analysis (PCA); a multivariate statistical approach used under Factor Analysis to reduce many variables into a few latent variables which are seen as capable of explaining the variability. The approach was effected under both conditions of presence and absence of metabolic risk. Other data settings were: within males, within females, in the rural and in urban communities.

Out of 977 participants, 17.4 %(n = 170), 11 %(n = 107), and 7.7 % (n = 75) had type 2 diabetes mellitus (T2DM), MS, and CMR, respectively. Among those participants in the presence of metabolic risk, it is interesting to note that contrary to established information, Blood Glucose Metabolism Disordering was observed as factor 1 while obesity was observed as factor 2. Glucose fasting and post load glucose formed a factor with higher loadings than other latent factors. Other revelations have been noted under different settings and combinations of the data. The data was split to produce different data sets leading to a combination of different findings.

Gender seemed to have some noticeable influence on data split according to gender. With the exception of BMI, levels of the rest of the variables were significantly higher in the presence of T2DM than non-diabetics. There was a negative correlation between glucose types and BP in the absence of CMR. Under factor analysis for all, BP (factor 1) and triglycerides-HDL (factor 2) explained

55.4% of the total variance while under factor analysis for the MS group, triglycerides-HDL-C (factor 1), BP (factor 2), and abdominal obesity-dysglycemia (factor 3) explained 75.1% of the total variance explained. In the absence of CMR, glucose (factor 1) and obesity (factor 2) explained 48.1% of the total variance. In the presence of CMR, 3 factors (factor 1 = glucose, factor 2 = BP, and factor 3 = obesity) explained 73.4% of the total variance.

With regard to the application of confirmatory factor analysis, multiple regression analysis revealed strong relationships between dependent and most independent variables. In addition, factor analysis showed heavy loadings of between endogenous and exogenous variables. Of great interest, confirmatory factor analysis justifiably confirmed the results found under exploratory factor analysis.

The MS pathogenesis may be more glucose-centered than blood pressure and abdominal obesity-centered without considering the lipid-lipoprotein component. MS should be specifically defined by ethnic cut-offs of waist circumference among Bantu Africans.

The objective of this paper is to provide an application of Factor Analysis (specifically *EFA* and *CFA*) to determine Metabolic Syndrome Components on anthropometric data from Kinshasa Hinterland, The Democratic Republic of Congo.

Key Words: Multivariate analysis, Factor analysis, Principal components analysis, Cardio-metabolic risk, metabolic syndrome, Bantu Africans, Type 2 diabetes, Exploratory factor analysis, Confirmatory factor analysis.

Acknowledgements

First, I must thank the wonderful and great God for giving me the energy and patience to work to the end of this thesis. Without Him, a lot things will have gone wrong. I am so grateful to my supervisors Professor Yongsong Qin, Dr. James Ndege and the late Professor Tyler Joanne for their academic advice, intellectual and professional guidance throughout the various stages of my thesis. I benefited greatly on the many occasions of our interactions. Thank you my supervisors. I must whole heartedly thank and acknowledge the Head of the Department of Statistics for his relentless and persistent contributions geared towards the improvement of this thesis. Thank you very much Dr. James Ndege and may the wonderful God bless you abundantly.

I extend my sincere gratitude to Doctor Geoff Howes for his continued support and advice. I thank you doctor for your suggestions towards the improvement of this thesis. Your experience was immensely resourceful. This came in handy and fulfilled a strong professional desire since the thesis was based on medical data. Issues of metabolic syndrome factors were clarified through your help. I am greatly indebted to Professor Benjamin Longombenza for his continued encouragement and persistent follow-ups with data acquisition and the subsequent analysis.

Many thanks go to the LOMO Clinic research team in the Democratic republic of Congo for making the availability of my thesis data a possibility. Thank you Pappy Nge and those other friends whom you worked with. I am humbled by the University of Fort Hare in general and the Department of Statistics in particular for giving me the opportunity to do my further studies and become a properly informed individual in society. This will go a long way in making me a proper academician.

Finally, I must thank my sons Dr Mark Wopicho Nasila and Mr. Jason Biketi Nasila for their wonderful professional advice. Jason, you have been so instrumental as if you thought that your father's success was your success.

Table of Contents

The Research Outline

Introduction

This chapter outlines the contents detailed under different chapters of this thesis. The researcher, under this chapter, gives brief settings of what has been written under the individual chapters. This thesis contains six chapters. The chapters which have been included in this project are stated and numerically identified as; Chapter one… Chapter six.

Chapter 1

This chapter discusses the introduction and the structure of the thesis. It also discusses the problem statement, the significance of the study as well as a brief introduction about the literature review and the possible analytical procedures to be used. The Chapter also states the general and specific objectives of the study.

Chapter 2

Chapter two discusses the literature review of the study in detail. The chapter explores in detail the literature of research done on similar topics and other relevant work that is in progress with particular emphasis on Cardio-metabolic Risk factors. Specific interest was on the applications of Multivariate Technique to Medical Research Data; particularly the two approaches of factor analysis including Exploratory Factor Analysis and Confirmatory Factor Analysis. Examples of applications of data analysis of similar research and results obtained have been included

Chapter 3

Chapter three discusses the materials and methods that were used in this study. The study was a cross-sectional survey conducted between January, and April 2005, in Kinshasa Hinterland, in the Democratic Republic Congo. The variables included in this study were; Height, Waist Circumference, Body Mass Index, etc.

Chapter 4

This chapter discusses the research data, the origin of the data, and the type of the data and the area of the origin of the data. Another issue under this topic was the survey procedure used in the data collection. These surveys captured medical data on individuals who either suffered from factor(s) of Metabolic Syndrome or not. The interest was on the ability of factor analysis as means of generating non-lipidic components among Bantu Africans at a high risk of T2DM and MS, selecting variables for Principal Component Analysis and Description of the Study Area (Kinshasa Hinterlands). Different data types were discussed under this chapter. This included primary as well as derived variables.

Chapter 5

This chapter focuses on data presentation, analysis and interpretation of the research findings. Different scenes were considered for this research data analysis which have been explained under this chapter for the following categories; analysis among all men and all women (the general population), analyzed according to gender and according to Presence and Absence of Metabolic Risk and analysis according to area of residence (rural or urban). The results output were based on computations of Exploratory Factor Analysis and Confirmatory Factor Analysis.

Chapter 6

This chapter discusses final conclusions and recommendations with perspectives incorporated into this study.

List of Tables

List of Figures

List of Abbreviations

AAFNMU: Australian Food and Nutrition Monitoring Unit

AMOS: Analysis of Moment Structures

ATP: Adult Treatment Panel

BMI: Body Mass Index

CFA: Confirmatory Factor Analysis

CHD: Coronary Heart Disease

CMR: Cardio-metabolic Risk

COV(X): Covariance of (X)

CRP: C - reactive protein

CVD: Cardiovascular Disease

DBP: Diastolic Blood Pressure

DM: Diabetes Mellitus

DRC: Democratic Republic of Congo

EGIR: European Group for Insulin Resistance

FPG: Fasting Plasma Glucose

HDL-C: High Density Lipoprotein-Cholesterol

HOMA-IR: Homoeostasis Assessment for Insulin Resistance

IFD: International Federation of Diabetes

IOTF: International Obesity Task Force

MS: Metabolic Syndrome

mTBI: Mild traumatic brain injury

PAI: Plasminogen Activator Inhibitor

PCA:	Principal Component Analysis
PCFA:	Principal Component Factor Analysis
P-value:	Probability Value
SAS:	Statistical Analysis System
SBP:	Systolic Blood Pressure
SD:	Standard Deviation
SPSS:	Statistical Package for Service Solutions
T2DM:	Type Two Diabetes Mellitus
TG:	Triglycerides
USA:	United States of America
WC:	Waist Circumference
WHO:	World Health Organization
PPCS:	Persistent Post-Concussive Syndrome

Chapter 1

1. Introduction

This chapter introduces the academic research study of the topic of the utilization of Factor Analysis to determine the clustering (component construction) of data that composed of variables on anthropometric measurements. Simply put, this research endeavored to respond to issues related to clustering of Metabolic Syndrome data.

1.1. Brief Statistical Introduction

Statistics is a branch of applied mathematics concerned with collecting, organizing, analyzing and interpreting numerical data (Rice, 2010). Statistics is also the mathematical study of the likelihood and probability of events occurring based on known quantitative data or a collection of data. Statistics, thus attempts to infer the properties of a large collection of data from inspection of a sample of the collection thereby allowing educated guesses to be made with a minimum of expense. There are two branches of Statistics which are stated and defined separately as follows:

- **Univariate statistics***:* includes all statistical techniques for analyzing a single variable of interest or a single dependent variable.
- **Multivariate statistics***:* includes all statistical techniques for analyzing two or more variables of interest or, two or more dependent variables. (Gupta and Kapoor, 2000)

This thesis is based on the application of two multivariate approaches to a health related profession, since the analysis carried out involves data that has many variables. To be more specific, the multivariate method used is factor analysis. Historically factor analysis was used primarily by psychology and education; however its use within the health science sector has become much more common during the past two decades. Factor analysis is commonly used in psychology, education, and more recently in the health-related professions (Willaims, Onsman and Brown, 2012).

The two factor analysis methods covered on this paper are Explanatory Factor analysis (EFA) and Confirmatory Factor Analysis (CFA). This methods are data reduction which infer presence of latent factors which are responsible for the shared variance in a set of observed variables/ items. Exploratory Factor Analysis is by definition 'exploratory' – the user does not specify a structure, and assumes each item/ variable could be related to each latent factor, whereas, CFA signifying 'confirmatory' – the user defines which observed variables/ items are related to the specified constructs or latent factors based on a priori theory or the results of EFA.

The two outlined approaches (CFA and EFA) were adopted to determine the clustering (component construction) of the data obtained. The results obtained show well-structured components using combinations of all or some of the following; eigenvalues, scree plot, total variance explained and the three dimensional figures of the Component Plot in Rotated Space using the verimax rotation procedure.

In summary, this study intends to respond to questions surrounding the philosophy of this research and further respond to "why" this research was conducted.

1.2. Background on Metabolic Syndrome

According to (Robert et al., 2013), (Huang, 2005) and (Okafor, 2012), Cardiovascular Research Center, Massachusetts General Hospital, Metabolic Syndrome refers to the co-occurrence of several known cardiovascular risk factors, including insulin resistance, obesity, atherogenic dyslipidemia and hypertension.

These conditions are interrelated and share underlying mediators, mechanisms and pathways. There has been recent controversy about its definition and its utility. In this study, a review of the current definitions for the metabolic syndrome and why the concept is important, is given. It identifies a subgroup of patients with shared pathophysiology who are at high risks of developing cardiovascular diseases and type 2 diabetes. By considering the central features of metabolic syndrome and how they are related, we may better understand the underlying pathophysiology and disease pathogenesis, Roberts et al. (2013).

Physicians and scientists have long known that certain conditions increase a person's risk of developing atherosclerotic cardiovascular disease (CVD). These risk factors include the family history of premature coronary disease, hypertension, hyperlipidemia, diabetes and smoking. It has also been noted further that age increases the risk of CVD, as does male gender and post-menopausal hormonal status in women. Of these risks, some environmental factors can be modified. However, genetic predisposition cannot be modified. The risk of CVD can be decreased by addressing these individual risk factors, both by lifestyle modifications and, if appropriate, pharmacological treatments may be instituted (National Cholesterol Education Program, 2002).

It has become increasingly clear that certain CVD risks tend to cluster, or occur together. Furthermore, the lifestyle modifications of dietary change and increased physical activity can significantly affect several risk factors simultaneously and, in so doing, reduce the risk of CVD. The clustering of some risk factors and their shared responsiveness to lifestyle modifications suggest that they are not independent of one another and that they share underlying causes, mechanisms and features (Grundy et al., 2005) and (Kahn et al., 2005).

The metabolic syndrome is a clustering of hyperglycemia/insulin resistance, obesity and dyslipidemia (National Cholesterol Education Program (NCEP 2001; Alberti and Zimmet, 2005; EGIR 2002). This is important for several reasons. Firstly, it identifies patients who are at high risk of developing atherosclerotic CVD and type 2 diabetes (T2D). Secondly, by considering the relationships between the components of metabolic syndrome, it is possible to better understand the pathophysiology that links them with each other and with the increased risk of CVD. Thirdly, it facilitates epidemiological and clinical studies of pharmacological, lifestyle and preventive treatment approaches.

1.3. Statement of the Research Problem

In sub-Saharan Africa, obesity, dyslipidemia, DM, hypertension and DM associated with hypertension, are emerging with cardiovascular complications because of change of lifestyles resulting from association with urbanization, migration, epidemiological transition, demographic transition, and nutritional transition (EGIR 2002, NCEP 2001, (Longo-Mbenza et al., 2011a), (Longo-Mbenza et al., 1999b), (Longo-Mbenza et al., 2010c), (Longo-Mbenza et al., 2010d), (Motala et al., 2003), (Kelliny et al., 2008), (Kasiam et al. 2008) (Longo-Mbenza et al., 2010) and (Kasiam et al., 2009).

Identifying patterns of African hypertensive diabetics at the primary care level can explain, at least in part, the differences observed in the prevalence or incidence of MS and cardiovascular diseases between different populations (Longo-Mbenza et al., 2011) and (Aljefree and Ahamed, 2014). Metabolic Syndrome (MS) is, to be more specific, a constellation (clustering) of six components (factors) identified as obesity, lipid-lipoprotein (fats) disorders, and increase in glucose (sugar), hypertension (blood pressure),inflammation/hypercoagulability (Jesmin et al., 2012), (Papanastasiou, 2013), (Okafor, 2012) and (Farook, 2015).

MS and other risk factors (smoking, physical inactivity, excessive alcohol intake, and inappropriate diet) determine high morbidity and mortality for cardiovascular disease CVD (heart attack, brain attack, peripheral vascular disease) or cardio-metabolic risk CMR (type 2 diabetes, kidney disease, retinopathy). Obesity, CVD, and CMR are emerging as epidemic conditions worldwide. Furthermore, lipid-lipoprotein profile is usually in normal range (triglycerides) or normal but very elevated high density lipoprotein-cholesterol (HDL-C) in black people in general (Jesmin et al., 2012) and (Papanastasiou, 2013).

However, Africa is not paying priority to early detection, treatment, prevention and control of atherosclerotic diseases (MS, CMR) from valid and reliable data (Kengne, Ntyintyane and Mayosi, 2012). Several statistical methods can be used to identify patterns of clustering in cardiovascular diseases such as T2DM and hypertension. One such important and useful technique is factor analysis – a multivariate technique used for treatment of large

multivariable data (Gurka et al., 2013), (Furguson et al., 2009), (Woolston et al., 2012), (Martinez-Vizcaino et al., 2010) and (Huo et al., 2013).

Indeed**,** Factor analysis is a statistical method used to describe variability among observed variables in terms of a potentially lower number of unobserved variables called factors (components) (Gurka et al. 2013), (Furguson et al. 2009), (Woolston et al., 2012), (Martinez-Vizcaino et al., 2010) and (Huo et al., 2013) . This includes two types of implementation, EFA and CFA. In other words, it is possible, for example, that variations in three or four observed variables mainly reflect the variations in a single unobserved variable, or in a reduced number of unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modelled as linear combinations of the potential factors, plus "error" terms (for CFA).

In general, Bantu Africans with and without adulthood diabetes mellitus (DM) and/or MS may be ethnically characterized by a particular clustering of components of MS. For that reason, the objective of this study was to provide a step-by-step description of the application of factor analysis and interpretations of the results based on anthropometric parameters; blood pressure and plasma glucose in the general population, men, women, rural and urban inhabitants and different types of DM. Thus, the detection of MS at early stages has been one of the objectives of this study and the researcher believes that without this, the problem of MS may never be addressed sufficiently.

So far, there is no known valid ethnic and gender specific pathophysiological mechanisms of; (why, when, where and how metabolic syndrome is produced or developed) affecting a class of Bantu people in Central Africa (Kaduka et al., 2012). It is the existence of this credibility gap that necessitated the present study. A comprehensive definition for MS and its key features would facilitate research into its causes and hopefully lead to new insights into pharmacological and lifestyle treatment approaches.

1.4. More on Problem Statement

The aim of this thesis was, among other analyses, to explore and confirm that Factor Analysis can optimally predict Metabolic Syndrome (Cardiometabolic Risk) in a Bantu setting using data collected in the general population in Kinshasa Hinterland in the Democratic

Republic of Congo. Furthermore, it was believed that this research will generate new knowledge on the general understanding of the MS as far as enlightening the general population with regard to the condition of MS.

1.5. Specific Objectives of the Study

The objectives of this study were stated as follows:

- To provide a step-by-step description of the application of factor analysis and interpretation of the results based on anthropometric parameters among Bantu Africans with different numbers and cut-offs of components of metabolic syndrome (MS);
- To describe the study population by general characteristics;
- To use exploratory and confirmatory factor analyses on a multiple of variables under different setups including: among all, by gender, by residence, on both traditional and cardio-metabolic risk;
- To correctly interpret the results obtained through factor analysis;
- To estimate the frequencies and proportions of T2DM, MS, and CMR;
- To ascertain the economic implications of the Metabolic Syndrome in the target population as well as the rest of the world.
- To compare Communalities within Factor Analysis under different rotation methods
- To compare Factor Analysis with other multivariate methods

1.6. Research Questions

This thesis was organized according to the following main and sub-research questions:

1.6.1. Main Research Question

Can Factor Analysis be considered the most appropriate statistical technique (mathematical model) to be used for clustering of Metabolic Syndrome data collected on Congolese people?

1.6.2. Sub-research Questions

- Is Factor Analysis able to generate components of non-lipidic data and as well be used for prediction of metabolic syndrome effects?

- Are EFA and CFA compatible?
- How many components (factors or dimensions) were created based on the available data with its metabolic syndrome (cardio-metabolic risk factors) depending on the data setting?
- What are the loading structures and the corresponding total variances explained among all, among the men and among the women in the presence and absence of metabolic syndrome?
- How should the analysis output be correctly interpreted for proper utilization of the final results?

1.7. The Conceptual Framework of this Study

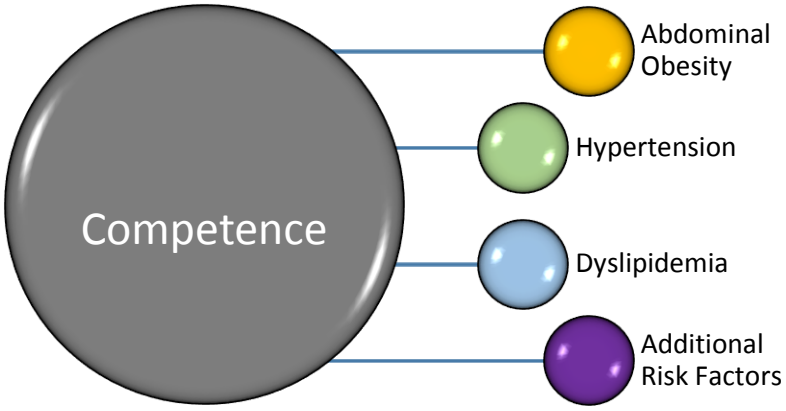

Figure 1 Conceptual model of Metabolic Syndrome using Factor Analysis, Roberts et al. (2013).

This process was based on different mathematical equations starting from the left (competence of factor analysis) to the right (components to be generated; Total-Abdominal-Peripheral Obesities, Hypertension and additional components) of the diagram (see Figure 1). Gender (males and females) was entered in the process as mediators.

The conceptual model of this thesis was a process built on a theoretical framework and on factor analysis in view of supporting the stated strategies in a variety of ways: By characterising clinical phenomena exploring relationships and discovering constructs that unite a set of metabolic risk factors using diagram (see Figure 1). The conceptual framework of this thesis is in relation to factor analysis defining a distinction between

exploratory and confirmatory models (Gurka et al. 2013), (Furguson et al., 2009), (Woolston et al., 2012), (Martinez-Vizcaino et al., 2010) and (Huo et al., 2013).

Factor Analysis is based on a matrix and is the same across all, based on gender and based on metabolic syndrome status. The manner in which the MS factors (components) correlate with this factor differs across the groups in a meaningful way. There exist gender specific scores by EFA and CFA better understood from conventional metabolic syndrome risk factors. Confirmatory factor analysis (CFA) is another procedure of evaluating the factor structures of MS based on the theoretical foundations set by EFA. Explanatory variables included anthropometric parameters; blood pressure, lipid-protein profile and glucose tolerance which were considered potential components when analysed for clustering under the following groups, All patients, T2DM Patients and Presence and absence of MS/CMR, using factor analysis.

It was reckoned that MS was characterising participants at a high risk of CVD according to the variances generated by intermediate mechanisms (Image factoring, Extraction sum of squared loadings, Determining the number of factors using Cattel rotations).

CFA was assumed more valid than EFA through multivariate normality, steps, and model fit. Consequently, Confirmatory Factor Analysis was effected on the data to confirm the findings already obtained at the exploratory analysis stage. Furthermore, CFA could lead the researcher to additional discoveries which could improve the final product of the output of the data analysis.

1.8. Rationale of the Study

This study was conducted to understand whether Factor Analysis in its totality could be an appropriate statistical technique to properly analyse medical data such as has been defined under this study. The emergency of serious medical problems associated with Metabolic Syndrome is of significant concern here. Prediction thereof and creating medical, administrative and societal awareness is the ultimate goal to be achieved by this research.

The analysis was based on all approaches for a study at its level. These included:

- Simple descriptive analyses relevant to this study;
- Exploratory and confirmatory factor analyses including the required features such as Scree Plot, Component Plot in Rotated Space, Total Variance Explained, Simple Component Matrix, Rotated Component Matrix, etc.;
- Significance tests.

The researcher was able to prove beyond any doubt, that Factor Analysis was a competent procedure of creating meaningful clusters when applied on these kinds of data. Very useful information was uncovered and revealed including creating factors such as Obesity, Fats, Sugar, and Blood Pressure which led to establishing clear medical markers for these data.

A very good and typical example was the component formed by Systolic Blood Pressure and Diastolic Blood Pressure. According to the analysis of the anthropometric data for this study, the two primary variables led to the creation and identification of the factor (component) identified as Blood Pressure.

This study was noted as an eye opener in the world of anthropometric research data analysis. The techniques utilized for data treatment for this study could be adopted for other bio-statistical data of the kind analysed and interpreted for this research. The results obtained by this study give unconditional early prediction of Cardio-metabolic Risk factors, promises contribution to national economies by reducing budget expenditures on the treatment of metabolic syndrome patients. At an international level, Metabolic Syndrome has caused untimely deaths, led to cursed amputations, led to a significant number of orphans. Furthermore, early predictions of factors associated with Metabolic Syndrome, is something which medical practitioners hope to achieve.

On the economic platform, governments and individuals have suffered heavy economic losses either directly or indirectly by spending millions on the management and treatment of sicknesses resulting from Metabolic Syndrome.

1.9. Significance of the Study

The findings of this thesis will impact on the understanding of the advantages and the limitations of factor analysis by statisticians, the organization of guidelines for the

prevention, early diagnosis and treatment, and control of T2DM, MS, CMR and CVD within well informed health care systems, health personnel, governments, patients and the general population.

Other beneficiaries of the results of this study will be individual medical practitioners, Departments of Health and its officials, Department of Education, universities, medical researchers, the World Health organization (WHO) and other interested both private and government individuals. We can imagine many medical researchers receiving useful information from the publication of this thesis and distribution to the university library as they undertake their research work.

Chapter 2

2. Literature Review

The present thesis provides information on the strategy for literature search to review previous publications on the applications of both exploratory and confirmatory factor analyses.

This chapter presented the views of various authors on this subject and evaluated them in the context of this thesis strategy. Summaries of this thesis were also written for the methodologies applied and for noted short comings.

The chapter has the objective of providing the literature review on the research subject of Metabolic Syndrome and the application of factor analysis on the data. The chapter was devoted to explaining to the reader the different terms used under this study and the review of literature which is stated as a survey of what has been written and published in relation to this study topic, its theory, or its research questions. It provided background for more extensive work.

2.1. Overview of the Factor Analysis Literature

Factor analysis uses mathematical procedures for the simplification of interrelated measures to discover patterns in a set of variables (Child, 2006). Attempting to discover the simplest method of interpretation of observed data is known as parsimony, and this is essentially the aim of factor analysis (Harman, 1976). Factor analysis has its origins in the early 1900's with Charles Spearman's interest in human ability and his development of the Two-Factor Theory; this eventually lead to a burgeoning of work on the theories and mathematical principles of factor analysis (Harman, 1976). The method involved using simulated data where the answers were already known to test factor analysis (Child, 2006). Factor analysis is used in many fields such as behavioral and social sciences, medicine, economics, and geography as a result of the technological advancements of computers.

Factor analysis is a widely used set of techniques in the behavioral sciences. It is also a primary technique for many researchers, especially those who conduct assessment-related

studies. The basic logic and mathematics of factor analysis were first described by Charles Spearman (1904b), and many variations of factor analysis were developed over the following century. Factor analysis is unique among multivariate statistical procedures in that it was developed mainly by psychologists in order to test hypotheses about the correspondence between scores on observed variables, or indicators, and hypothetical constructs (latent variables), or factors, presumed to affect those scores. Spearman and his contemporaries (e.g., Thomson, 1920) used factor analysis to evaluate models about the nature and organization of intelligence. Factor analysis is still widely used today in mental test studies as it is in many other research areas as a means to discover and identify latent variables, given initially only sample covariances among a set of indicators (Mulaik, 1987).

Factor analysis is among the most versatile and controversial techniques for analyzing data in the behavioral and social sciences. Factor analysis is commonly used to analyze complex data yet it is often misused and misinterpreted. For example, Gould's 1981 description of factor analysis is a popular treatment of the topic, yet Carroll (1995) criticized Gould's interpretation of factor analysis. This commentary introduces readers to general issues surrounding factor analysis and suggests some best practices when using and reporting results of factor analyses in gifted education. Interested readers should also consult technical treatments of the topic that provide step-by-step guidance, such as those provided by Pedhazur and Schmelkin (1991), Tabachnick and Fidell (2001), Hurley et al. (1997), Kieffer (1999), and Byrne (1998, 2001), among many others.

Over the years there have been thousands of published factor analytic studies (e.g., Costello & Osborne, 2005), so the impact of factor analysis in terms of sheer volume of the research literature is undeniable. Whether the typical factor analytic study also makes a substantive contribution, however, has been a matter of longstanding debate (e.g., Furfey & Daly, 1937). One challenge is that factor analysis has many decision points. This aspect of the technique is difficult for novices, who must navigate the analysis through a myriad of options about variable selection and sampling, the form of the input data, the method of factor extraction, and interpretive strategies, to name a few. A series of bad choices can compromise the results. It is also does not help that default options in some computer procedures for factor analysis are not actually the best choices in many studies. Based on

critical reviews of the use of factor analysis in several different research areas (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Watson & D. Thompson, 2006), it seems that many, if not most, factor analytic studies have at least one serious flaw. Common problems include sample sizes that are too small and failure to report sufficient numerical results so that the work can be critically evaluated.

There are two major classes of factor analysis: Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA). Broadly speaking EFA is heuristic. In EFA, the investigator has no expectations of the number or nature of the variables and as the title suggests, is exploratory in nature, (Graham, Guthrie, & Thompson, 2003). CFA attempts to confirm hypotheses and uses path analysis diagrams to represent variables and factors, whereas EFA tries to uncover complex patterns by exploring the dataset and testing predictions (Child, 2006). Factor analysis operates on the idea that measurable and observable variables can be reduced to fewer latent variables that share a common variance and are unobservable, which is known as reducing dimensionality (Bartholomew, Knott, & Moustaki, 2011). These un-observable factors are not directly measured but are essentially hypothetical constructs that are used to represent variables (Cattell, 1973).

EFA is used when a researcher wants to discover the number of factors influencing variables and to analyze which variables 'go together' (DeCoster, 1998). A basic hypothesis of EFA is that there are different common 'latent' factors to be discovered in the dataset, and the goal is to find the smallest number of common factors that will account for the correlations (McDonald, 1985). Another way to look at factor analysis is to call the dependent variables 'surface attributes' and the underlying structures (factors) 'internal attributes' (Tucker & MacCallum, 1997). Common factors are those that affect more than one of the surface attributes and specific factors are those which only affect a particular variable (Tucker & MacCallum, 1997).

Also, a determining factor is based on the assumption that there is a linear relationship between the factors and the variables when computing the correlations (Gorsuch, 1983). For something to be labeled as a factor it should have at least 3 variables, although this depends on the design of the study (Tabachnick & Fidell, 2007).

The recommended sample size is at least 300 participants, and the variables that are subjected to factor analysis each should have at least 5 to 10 observations (Comrey & Lee, 1992). We normally say that the ratio of respondents to variables should be at least 10:1 and that the factors are considered to be stable and to cross-validate with a ratio of 30:1. A larger sample size will diminish the error in your data and so EFA generally works better with larger sample sizes. However, Guadagnoli and Velicer (1988) proposed that if the dataset has several high factor loading scores (> .80), then a smaller small size (n > 150) should be sufficient. A factor loading for a variable is a measure of how much the variable contributes to the factor; thus, high factor loading scores indicate that the dimensions of the factors are better accounted for by the variables. Next, the correlation r must be .30 or greater since anything lower would suggest a really weak relationship between the variables (Tabachnick & Fidell, 2007). It is also recommended that a heterogeneous sample is used rather than a homogeneous sample as homogeneous samples lower the variance and factor loadings (Kline, 1994). Factor analysis is usually performed on ordinal or continuous variables, although it can also be performed on categorical and dichotomous variables. If a dataset contains missing values, one will have to consider the sample size and if the missing values occur at a nonrandom pattern. Generally speaking, cases with missing values are deleted to prevent overestimation (Tabachnick & Fidell, 2007). Finally, it is important that one checks for an absence of multicollinearity and singularity within your dataset by looking at the Squared Multiple Correlation (Tabachnick and Fidell, 2007). Variables that have issues with singularity (i.e., SMC close to 0) and multicollinearity (SMC close to 1.0) should be removed from the dataset.

Schmitt (2011) had an overview of the use of factor analysis where researchers used inappropriate sample size intended to achieve accurate parameter estimates and adequate power, a factor model and estimation method, a method for determining the number of factors and evaluating model fit, and a rotation criterion all of which were based on outdated methods. He castigated those researchers who decide to conduct a factor analytic study and evaluate a model, and advised it is important that they at least consider sample size, factor models and estimation methods, procedures for determining the number of factors and evaluating model fit, and rotation criteria. Researchers are encouraged to

explore further the references provided when more depth is required and/or desired on a particular topic. And though relatively current, researchers conducting factor analysis need to stay abreast of the current methods.

Knafl and Grey (2007) state that Medical research studies which utilize survey instruments consisting of responses to multiple items combined into one or more scales stand to benefit from methods for evaluating exploratory and confirmatory factor analysis models with decisions about covariance structure, including the number of factors, the factor extraction procedure, the allocation of survey items to summated scales and the extent of inter-scale dependence, made objectively using a likelihood-based form of cross-validation. This approach is demonstrated through example analyses using baseline data for three survey instruments from a clinical trial involving adolescents with type 1 diabetes. This paper showed that the Likelihood cross-validation provides an objective basis for assessing factor analysis models, for both exploratory as well as confirmatory purposes. It can be used to assess the impact of different numbers of factors, alternative factor extraction procedures, factor loading adjustments based on rotation schemes, comparison of possible summated scales, specifications of inter-scale correlation, realignment of item–scale allocation and item removal. It can be used for these purposes when assessing the applicability of established scales to specific data, by comparing alternative scales based on EFA factor scores or scales determined from possibly rotated versions of these factor scores to scales based on recommended item–scale allocations or on other theoretical considerations.

2.2. Exploratory Factor Analysis

Geometric Approach

Factor analysis can be examined through a geometrical approach to gain a better understanding of how the technique works. In a coordinate system, the factors are represented by the axes and the variables are lines or vectors (Cattell, 1973).

Limitation of EFA

One of the limitations of this technique is that naming the factors can be problematic. The variables form the data may correlate with each another to produce a factor despite having

little underlying meaning for the factor (Tabachnick & Fidell, 2007). Tabachnick and Fidell (2007), further explained that it is not recommended to pool results from several samples or from the same sample at different points in time as these methods may obscure the findings. The limitations and special considerations required when performing factor analysis on categorical and dichotomous variables are beyond the scope of this paper. Refer to 'Recent Developments in the Factor Analysis of Categorical Variables' by Mislevy (1986) and 'Factor Analysis for Categorical Data' by Bartholomew (1980) for further explanation. The contribution of a variable is measured by a procedure called factor loadings (Kline, 2015).

The factor loadings give us an idea about how much the variable has contributed to the factor; the larger the factor loading the more the variable has contributed to that factor (Harman, 1976). Factor loadings are very similar to weights in multiple regression analysis, and they represent the strength of the correlation between the variable and the factor (Kline, 1994).

Variance

Factor analysis uses variances to produce communalities between variables. The variance is equal to the square of the factor loadings (Child, 2006). In many methods of factor analysis, the goal of extraction is to remove as much common variance in the first factor as much as possible (Child, 2006). The communality is the variance in the observed variables which are accounted for by a common factor or common variance (Child, 2006). The communality is denoted by h^2 and is the summation of the squared correlations of the variable with the factors (Cattell, 1973). A particular set of factors is said to explain a lot of the variance of a variable if it has a high communality (Kline, 1994). Often times variables with low communalities (less than .20 so that 80% is unique variance) are eliminated from the analysis since the aim of factor analysis is to try and explain the variance through the common factors (Child, 2006). Factors may be uncorrelated or correlated with each other (Harman, 1976). Generally, the cumulative percentage of variance is extracted after each factor is removed from the matrix, and this cycle continues until approximately 75-85% of the variance is accounted for (Gorsuch, 1983). The percentage variance tells us how much each factor contributed to the total variance.

Factor Extraction

Factor analysis is based on the 'common factor model' which is a theoretical model. This model postulates that observed measures are affected by underlying common factors and unique factors, and the correlation patterns need to be determined. There is an array of extraction methods available, but we will briefly touch on a few commonly used techniques that are available on SPSS. Maximum Likelihood attempts to analyze the maximum likelihood of sampling the observed correlation matrix (Tabachnick & Fidell, 2007). A useful summary of extraction methods can be found in Table 13.7 (p. 633) in 'Using Multivariate Statistics (5th ed.)' by Tabachnick and Fidell (2007).

Factors are then extracted successively until there is a large enough of variance accounted for in the correlation matrix (Tucker & MacCallum, 1997). Principal Axis Factoring is recommended when the data violates the assumption of multivariate normality (Costello & Osborne, 2005). Principal Component Analysis is used to extract maximum variance from the data set with each component thus reducing a large number of variables into smaller number of components (Tabachnick & Fidell, 2007). Principal Component Analysis is a data reduction technique and the issues of whether it is truly a factor analysis technique has been raised (Costello & Osborne, 2005). That is, Principal Components produces components whereas Principal Axis Factoring produces factors. There are also differences in how the correlation matrix is constructed and how the communalities are calculated when comparing these techniques (Kline, 1994; Tucker & MacCallum, 1997). Overall, the factor loadings are fairly similar and you will need to perform rotation regardless of the extraction technique (Tabachnick & Fidell, 2007).

Rotational Methods

Factors are rotated for better interpretation since un-rotated factors are ambiguous. The goal of rotation is to attain an optimal simple structure which attempts to have each variable load on as few factors as possible, but maximizes the number of high loadings on each variable (Rummel, 1970).

Ultimately, the simple structure attempts to have each factor define a distinct cluster of interrelated variables so that interpretation is easier (Cattell, 1973). Orthogonal rotation is when the factors are rotated 90° from each other, and it is assumed that the factors are uncorrelated (DeCoster, 1998; Rummel, 1970). Costello & Osborne, (2005) argue that this is less realistic since factors generally are correlated with each other to some degree. A summary of the rotation techniques can be found in Table 13.9 (p. 639) in 'Using Multivariate Statistics (5th ed.)' by Tabachnick and Fidell (2007). Varimax rotation. Quartimax involves the minimization of the number of factors needed to explain each variable (Gorsuch, 1983). Oblique rotation is more complex than orthogonal rotation, since it can involve one of two coordinate systems: a system of primary axes or a system of reference axes (Rummel, 1970). Direct Oblimin attempts to simplify the structure and the mathematics of the output, while Promax is expedient because of its speed in larger datasets. Promax involves raising the loadings to a power of four which ultimately results in greater correlations among the factors and achieves a simple structure (Gorsuch, 1983)

Interpretations of Factor Loadings

When interpreting the factors, one needs to look at the loadings to determine the strength of the relationships. Factors can be identified by the largest loadings, but it is also important to examine the zero and low loadings in order to confirm the identification of the factors (Gorsuch, 1983). There should be few item crossloadings (i.e., split loadings) so that each factor defines a distinct cluster of interrelated variables. A crossloading is when an item loads at .32 or higher on two or more factors (Costello & Osborne, 2005). The signs of the loadings show the direction of the correlation and do not affect the interpretation of the magnitude of the factor loading or the number of factors to retain (Kline, 1994). A general rule to determine the reliability of the factor is to look at the relationship between the individual rotated factor loading and the magnitude of the absolute sample size. That is, the larger the sample size, smaller loadings are allowed for a factor to be considered significant (Stevens, 2002). According to a rule of thumb, using an alpha level of 0.01 (two-tailed), a rotated factor loading for a sample size of at least 300 would need to be at least 0.32 to be considered statistically meaningful (Tabachnick & Fidell, 2007).

Number of Factors to Retain

Extracting too many factors may present undesirable error variance but extracting too few factors might leave out valuable common variance. One criterion that can be used to determine the number of factors to retain is Kaiser's criterion which is a rule of thumb. This criterion suggests retaining all factors that are above the eigenvalue of 1 (Kaiser, 1960). Another criterion is based on Jolliffe's criterion which recommends retaining factors above .70 (Jolliffe, 1986). It has been argued that both criteria may result in overestimation in the number of factors extracted (Costello & Osborne, 2005; Field, 2009); therefore, it is suggested to use the scree test in conjunction with the eigenvalues to determine the number of factors to retain. The scree test (*see Figure 3*) consists of eigenvalues and factors (Cattell, 1978). The scree test is only reliable when you have a sample size of at least 200. In situations when the scree test is hard to interpret (e.g., clustered data points at the point of inflexion), you will need to rerun the analysis several times and manually set the number of factors to extract each time (Costello & Osborne, 2005).

Factor Scores

A factor score can be considered to be a variable describing how much an individual would score on a factor. One of the methods to produce factor score is called Bartlett method (or regression approach) which produces unbiased scores that are correlated only with their own factor. Another method is called the Anderson-Rubin method which produces scores that are uncorrelated and standardized. The method that you choose will depend on your research question, but the Bartlett method is the most easily understood (Tabachnick & Fidell, 2007). Factor scores can be treated as variables for further statistical analyses of variables (e.g., ANOVA) or can be used to overcome the issue of multicollinearity as uncorrelated variables can be produced.

2.3. Confirmatory Factor Analysis (CFA)

CFA is a special case of the structural equation model (SEM), also known as the covariance structure (McDonald, 1978) or the linear structural relationship (LISREL) model (Jöreskog & Sörbom, 2004). SEM consists of two components: a measurement model linking a set of observed variables to a usually smaller set of latent variables and a structural model linking

the latent variables through a series of recursive and non-recursive relationships. CFA corresponds to the measurement model of SEM and as such is estimated using SEM software. Kline (2013) points out three steps involved in confirmatory Factor Analysis. Which are; specification and Identification, Estimation and Goodness of Fit (GOF).

Specification and Identification

A CFA measurement model is identified if it is theoretically possible for the computer to derive a unique estimate of every model parameter (Kline, 2013). There are two necessary but insufficient requirements for identification: (1) Every factor and error term must be assigned a scale, and (2) the model degrees of freedom must be at least zero ($df_M > 0$) (Kline, 2010). In particular, specifying that an indicator depends on more than a single factor or that a pair of error terms is correlated is possible only if certain additional requirements are met (DeCoster, 1998). These extra requirements are summarized in the form of identification heuristics for determining whether a nonstandard model is identified (e.g., Kenny, Kashy, & Bolger, 1998; Kline, 2010, Chapter 6), but these heuristics are not always straightforward to apply for complex models with multiple correlated errors or indicators with ≥2 pattern coefficients.

Estimation

The default method in CFA is Maximum Likelihood (ML), which in SEM analyzes covariance matrices only and simultaneously estimates all model parameters in an iterative algorithm (Kline, 2013). Computer procedures for ML estimation often begin iterative estimation by generating default initial estimates of certain model parameters known as start values. For example, the EQS program for SEM (including CFA) (Bentler, 2006) assumes in the first iteration that all un-standardized pattern coefficients in CFA models equal 1.0. Kline argues that the method of ML estimation assumes multivariate normality, and the method is not robust against violations of this assumption. This means that it is necessary to carefully screen the raw data and deal with problems, such as extreme outlier scores or severely non-normal univariate distributions that contribute to multivariate non-normality. Kline (2010, chapter 3) describes how to screen the data and prepare a "proper" matrix summary for ML estimation.

Goodness of Fit

There are two main classes of statistics in SEM that evaluate the correspondence between model and data, model test statistics and approximate fit indexes. The most widely reported test statistic is the model chi-square, χ_M^2, with degrees of freedom that equal df_M, the model degrees of freedom (Klein, 2010). The statistic χ_M^2 assumes multivariate normality, which is also required in ML estimation.

In small samples, it can happen that the power of the model chi-square test is so low that it is difficult to correctly reject a false model (e.g., MacCallum et al., 1996). In very large samples, it can happen that χ_M^2 is statistically significant even though the magnitudes of model-data discrepancies are slight. For this reason, researchers in the past tended to ignore the results of the model chi-square test even in samples that were not very large. However, this practice is now viewed by more and more methodologists as excessively lax (e.g., Barrett, 2007). Likewise, a model chi-square test result that is not statistically significant does not automatically lead to the decision to retain the model, especially if the sample size is not very large. Further evaluation of a model's fit to the data is also needed in this case (Kline, 2010).

2.4. Metabolic Syndrome Literature Review

The term Metabolic Syndrome is a subject that has received much attention in the recent times, due to increasing awareness of its association with cardiovascular morbidity and human morbidity leading to high death rates. However, it is a concept that dates back to over 5 decades. Its existence was first observed as clustering of hypertension, hyperglycemia, and gout as described by Kylin in the 1920s. Later, Jean Vague in 1947 noted its mystic association with android obesity (Okafor, 2012).

The effect of hypo-caloric, low-carbohydrate diet on obese patients with diabetes, hypercholesterolemia, and hypertriglyceridemia was later reported by Avogadro, Crepaldi, and co-workers. Other important historical developments include the use of the term "metabolic syndrome" by (Haller, 1977) to describe the associations of obesity, diabetes mellitus, hypolipoproteinemia, hyperuricemia, and hepatic steatosis when highlighting the additive effects of risk factors on atherosclerosis.

Singer in 1977 also used this term for the associations of obesity, gout, diabetes mellitus, and hypertension with hypolipoproteinemia. In 1977 and 1978, Phillips developed the concept that risk factors for myocardial infarction form a "constellation of abnormalities" that is associated not only with heart disease but also with aging, obesity, and other clinical conditions. These abnormalities included glucose intolerance, hyperinsulinemia, hypercholesterolemia, hypertriglyceridemia, and hypertension. In 1988, in his Banting lecture, Reaven proposed insulin resistance as the underlying factor and named the constellation of abnormalities, "Syndrome". This however did not include abdominal obesity. The syndrome has been variously called "The Deadly Quartet" by Kaplan and The Insulin Resistance Syndrome. Notwithstanding the various aforementioned evolutionary nomenclatures, the term "metabolic syndrome" has remained widely accepted, and currently is the term that is in use globally (Okafor, 2012).

2.5. Definitions of Metabolic Syndrome

Metabolic syndrome is a constellation of interrelated risk factors of metabolic origin (metabolic risk factors) that appear to directly promote the development of atherosclerotic cardiovascular disease (NCEP, 2001), (Alberti and Zimmet, 2005), (EGIR, 2002) and (Nasila et al., 2013). The primary components of metabolic syndrome included insulin resistance, obesity, dyslipidemia, and hypertension but with continuing emergence of new facts, the list tends to be growing. The components therefore now include insulin resistance, hyperinsulinemia, central obesity, hypertension, dyslipidemia (increase in plasma triglycerides (TG)), decrease in high density lipoprotein cholesterol (HDL-C), an LDL particle pattern shifted to small dense particles (type B pattern), pro-coagulant state (increased plasma fibrinogen, increased plasminogen activator inhibitor-1 (PAI-1)), vascular abnormalities (increase in urinary albumin excretion, endothelial dysfunction), inflammatory markers, and hyperuricemia (Roberts et al., 2014).

The wide interest generated by this subject has resulted in several definitions being put forward by different expert groups. These definitions indicate attempts that have been made to describe the syndrome from various perspectives of the expert groups but despite all these, no unified definition has been adopted. The quest to fill this gap for a unified worldwide definition resulted in the definition proposed by the International Diabetes

Federation (IDF). This was the outcome of a consensus workshop held from May 12 to 14 2004 in London, UK. The IDF also realized that there are still some unanswered questions; hence areas for future research to help describe the syndrome better were identified (Okafor et al., 2014). These areas for future research can be grouped into epidemiological, clinical, and biochemical characterizations. Due to its continued evolution and other reasons, debate is still ongoing in some quarters about the usefulness of the syndrome (Okafor et al., 2014).

Current Discussions on Metabolic Syndrome

International Organizations NCEP (2001), the European Group for the Study of Insulin Resistance EGIR (2002) define the Metabolic Syndrome (Cardio-metabolic Risk) by the presence of three or more of the following traditional risk factors:

- Abdominal Obesity, the first component which is defined by the elevation of Waist Circumference (WC);
- Hypertriglyceridemia;
- Low – high density lipoprotein Cholesterol or (HDL-C);
- Hyperglycemia (elevated sugar-glucose);
- Hypertension (elevated blood pressure). Albert KG Circulation 2009; 120:1640-1645.

These definitions of the stated traditional risk factors explain the existing controversy about the definition and classification of Metabolic Syndrome worldwide (Kahn et al., 2005). Additional risk factors of Metabolic Syndrome have been identified as follows (Anna et al., 2011) in Stanford Center for Research on Disease Prevention, Stanford University School of Medicine, Stanford, California);

- Biomarkers of subclinical inflammation;
- Markers of the liver function;
- Serum Uric Acid;
- Serum and urinary albumin.

The World Health Organization (WHO) first developed its definition in 2005. Because insulin resistance was felt to be central to the pathophysiology of Metabolic Syndrome, evidence for

insulin resistance is an absolute requirement in the WHO's definition. This could be impaired fasting glucose [IFG, defined as a fasting glucose level above the predetermined cutoff; commonly 100 milligrams per deciliter (mg/dl)] or impaired glucose tolerance (IGT, defined as a glucose level above a predetermined cutoff, commonly 140 mg/dl, for 120 minutes after ingestion of 75 grams of glucose load during an oral glucose tolerance test). Alternatively, other measures could serve as evidence of insulin resistance, such as an elevated homeostatic model assessment of insulin resistance (HOMA-IR) value, which is proportional to the product of the fasting insulin and fasting glucose level. Finally, euglycemic hyperinsulinemic clamp studies could be used as evidence of insulin resistance (Charmane et al., 2012). It continues "in addition to this absolute requirement for insulin resistance, two additional criteria have to be met. These include obesity, dyslipidemia, hypertension and microalbuminuria".

2.6. Statistical Results of a Study on Korean Subjects

Results of a study carried out by (Song et al., 2015) have shown that low consumption of fruits and dairy foods is associated with Metabolic Syndrome. To confirm this hypothesis, a study was carried out on Korean adults from outpatient clinics to examine differences in nutrient intake and food consumption by the presence of metabolic syndrome. This study was performed on 668 subjects whose nutrient intake did not differ by the presence of Metabolic Syndrome in both men and women. Men with MS had lower consumption and percentage of the recommended allowance for fruits compared to those without MS (1.6 vs. 1.1 servings/day, $P-value$ = 0.001; 63.5 vs. 49.5%, $P-value$ = 0.013). Women with MS showed lower consumption and percentage of the recommended allowance for dairy foods than those without MS (0.8 vs. 0.5 servings/day, $P-value$ = 0.001; 78.6 vs. 48.9%, $P-value$ = 0.001).

Application of Exploratory Factor Analysis in Anthropometric Data in Jordan

Khader et al. (2011), conducted a study to explore the factor structure of the central metabolic syndrome variables in Jordanian children and adolescents using exploratory factor analysis to understand the factor structure and relative importance of the metabolic syndrome components. The study included 665 children and adolescents who were identified in a national population-based household survey in Jordan. Their anthropometric

and laboratory measurements were obtained. Factor analysis was performed on standardized variables to produce the minimum number of factors that retains as much of the total variance in the original data as possible.

Factor analysis showed that one common factor is not sufficient to underlie metabolic syndrome. Four factors were extracted in the exploratory factor analysis; adiposity factor, blood pressure factor, lipids factor, and blood glucose factor. The cumulative percent of variance accounted for by the four factors together was 78.7% in male children, 86.9% in female children, 82.5% in male adolescents, and 83.4% in female adolescents. The adiposity factor accounted for the largest proportion of the total variance in the four groups.

The factor analysis of cardiovascular risk clustering among Jordanian children and adolescents suggested that multiple factors account for the clustering of the metabolic syndrome component. Obesity accounts for the maximum variance in clustering and appears to be a more powerful correlate of cardiovascular risk in children and adolescents (Khader et al., 2011).

FA of Persistent Post-Concussive Symptoms within a Military Sample with Blast Exposure

A study was conducted by (Franke et al., 2015) on 181 service members and veterans with at least one significant exposure to a blast during deployment within the two years prior to study enrollment. The objective was to determine the factor structure of persistent post-concussive syndrome (PPCS) symptoms in a blast-exposed military sample and validate factors against objective and symptom measures. The setting was the Veterans Affairs Medical Center and military bases. The study data analyses were Confirmatory and exploratory factor analysis of the Rivermead Post-concussion Questionnaire (RPQ). Measurements were taken on; RPQ, PTSD Symptom Checklist-Civilian, Center for Epidemiologic Studies Depression inventory, Sensory Organization Test, Paced Auditory Serial Addition Test, California Verbal Learning Test, Delis-Kaplan Executive Function System subtests.

The three-factor structure of PPCS was not confirmed. However, a four-factor structure was extracted, and factors were interpreted as reflecting; emotional, cognitive, visual, and

vestibular functions. All factors were associated with scores on psychological symptom inventories; visual and vestibular factors were also associated with balance performance. There was no significant association between the cognitive factor and neuropsychological performance, nor between a history of mTBI and factor scores.

It was concluded that Persistent post-concussive symptoms observed months after blast exposure seemed to be related to four distinct forms of distress, but not to mTBI per se, with vestibular and visual factors possibly related to injury of sensory organs by blast.

This was another demonstration of the capability of Factor Analysis in a complicated and rare practical setting.

2.7. Metabolic Syndrome and Gender

Gender-specific differences have been demonstrated by different workers. Metabolic syndrome appears to be more common in females like obesity whereas hypertension appears to be more common in males. The prevalence of metabolic syndrome was only observed to be higher among males from the Jos plateau of Nigeria where the authors noted that the high activity profile of women may have contributed to this observation. This pattern is at variance with the findings from the north-western Nigeria (Sokoto) where the religious practice of putting the women in Purdah makes them sedentary (Krumsiek et al., 2015), (Kim et al., 2010), (Liu et al., 2012) and (Regitz-Zagrosek, 2012).

Age-adjusted prevalence of central obesity (using NCEP-ATP III and IDF definitions, based on waist circumference) was found to be higher in women compared to men and were lower in the rural than in the urban areas. Among Cameroonians, considering those with two components of metabolic syndrome, the most frequent combination was central obesity and high blood pressure, which was more predominant among women than men (81% vs. 52%).

Combinations of high blood pressure and hypercholesterolemia (24% vs. 6%) and high blood pressure and hyperglycaemia (12% vs. 6%) demonstrated male predominance. In South Africa, greater incidence of risk factors for the metabolic syndrome occurred in males though obesity was more common in females (25% vs. 14%). Both genders had abnormally

high mean TG but male predominance appeared to be observed for dyslipidemia. Metabolic syndrome was seen to be more common in males in Jos, Nigeria (Chukwuonye et al., 2013) and (Iloh et al., 2013).

2.8. Metabolic Syndrome and Age

Metabolic Syndrome, predicting Type 2 Diabetes Mellitus (T2DM) and Cardiovascular Disease (CVD), is no longer an illness for ageing people (Anna et al., 2011). Metabolic syndrome was initially recognized as an adult disorder because its early descriptions were made in adults and its constituent components are disorders which are seen common in adults or they are associated with aging (elderly). The currently emerging data in Africa have mainly been from young-adult populations (>20 years) though some studies had involved subjects that were aged <20 years (Okafor, 2012). Several workers have observed that the prevalence of metabolic syndrome increases with age. Prevalence increased from 11% in subjects aged 20-29 years to 89% in those aged 70-79 years in Nigeria. Available data suggest that despite the increasing trend in prevalence as age increases, adults who may be classified as middle aged (40-60 years) are predominantly affected.

This trend is also recorded in the mean ages of around 50 years in these studies (Bosu, 2015), and (Ugwuja et al., 2013). Metabolic syndrome today is not only reported in adults but it is now beginning to occur in children and adolescents due to the growing obesity epidemic within this young population. As a result, the IDF has also developed a consensus definition for identifying this syndrome among the young. The intention is "to obtain a universally accepted tool which is easy to use for the early diagnosis of metabolic syndrome, in order to take preventive measures before the child or adolescent develops diabetes or cardiovascular disease. International Diabetes Federation has given a consensus definition of the metabolic syndrome in children and adolescents (Wang et al., 2013) and (Ugwuja et al., 2013).

Approximately 22 million children under the age of 5 years were estimated to be overweight or obese by the World Health Organization (WHO, 2004) while according to the International Obesity Task Force (IOTF), at least 10% of school-aged children between 5 and 17 years are overweight or obese. In the United States of America for instance, the rate

of overweight and obesity among children and adolescents aged 6-18 years increased from 15% in the 1970s to more than 25% in the 1990s.

Such increases are not restricted to developed countries; many low and middle-income countries are becoming largely involved. Globally, it is estimated that 17 million of the 22 million children under 5 years live in major economically developing countries. In China, for example, the rate of over-weight and obesity observed in a study of urban school children increased from almost 8% in 1991 to more than 12% after 6 years, while in Brazil, the rate of over-weight and obesity among children and adolescents 6-18 years old increased from 4% in the mid-1970s to over 13% in the late 1990s. A survey among South African undergraduate students revealed that metabolic risk factors were evident in this much younger population (60%) than was commonly expected with gender-specific differences being observed. This may not be unconnected with the epidemiologic transition being witnessed in many developing nations, with South Africa appearing to be advanced compared to other African nations (WHO, 2004).

In Egypt, metabolic syndrome was found in 7.4% of 4250 adolescents with nearly 25% having the full components of the syndrome among those with high values of different components. The odds of having the syndrome was increased by positive family history of obesity and diabetes mellitus. Similarly, pro inflammatory markers were also found to be common in subjects who participated in the study. In some nations of Africa as well as other developing nations, it has been shown that a high socio-economic status may be related to positive obesity status. This observation is in contrast to what is seen in developed nations where high educational status (one of the determinants of socioeconomic status) seems to be protective against metabolic syndrome; this benefit is attributed to potential mechanisms such as exposure to less psychosocial and material stress, better health knowledge, and better health behavior (Rodriguez et al., 2015) and (Zeba et al., 2012).

2.9. Metabolic Syndrome and Ethnicity

Literature reveals different ethnic-specific cutoffs for metabolic syndrome worldwide (Kassim et al., 2009), (Shai et al., 2006), (Deurenberg-Yap et al., 2000a), (Wen et al., 2009),

(Deurenberg-Yap et al., 2002b) and (Misra et al., 2009). The chance of developing diabetes, heart disease, and other weight-related health risks increases with increase in body mass index (BMI).There is, however, strong evidence that at any given BMI level, these health risks are markedly higher in some ethnic groups than others.

The Nurses' Health Study, for example, tracked patterns of weight gain and diabetes development in 78,000 U.S. women, to see if there were any differences by ethnic group. All women were healthy at the start of the study. After 20 years, researchers found that at the same BMI, Asians had more than double the risk of developing type 2 diabetes than whites; Hispanics and blacks also had higher risks of diabetes than whites, but to a lesser degree. Increases in weight over time were more harmful in Asians than in the other ethnic groups: For every 11 pounds Asians gained during adulthood, they had an 84 percent increase in their risk of type 2 diabetes; Hispanics, blacks, and whites who gained weight also had higher diabetes risks, but again, to a much lesser degree than Asians. Several other studies have found that at the same BMI, Asians have higher risks of hypertension and cardiovascular disease than their white European counterparts, and a higher risk of dying early from cardiovascular disease or any cause (Matthew et al., 2013), (Shai et al., 2006), (eurenberg-Yap et al., 2000_a) (Wen et al., 2009), (Deurenberg-Yap et al., 2002_b) and (Misra et al., 2009).

Researchers are still contemplating as to why Asians have higher weight-related disease risks at lower BMIs. One possible explanation is body fat. When compared to white Europeans of the same BMI, Asians have 3 to 5 percent higher total body fat.

(Alberti, Zimmet & Swaw, 2006) South Asians, in particular, have especially high levels of body fat and are more prone to developing abdominal obesity, which may account for their very high risk of type 2 diabetes and cardiovascular disease. (WHO/IDF Consultation, 2006; EGIR, 2002). In contrast, some studies have found that blacks have lower body fat and higher lean muscle mass than whites at the same BMI, and therefore, at the same BMI, may be at lower risk of obesity-related diseases. Keep in mind, though, that in the U.S., the prevalence of obesity is higher in non-Hispanic blacks than in non-Hispanic whites, so the

overall burden of obesity-related diseases is still higher in this group. Read more about obesity trends in the U.S. and other countries.

While genetic differences may be at the root of these different body fat patterns in Asians and other ethnic groups, environmental factors seem to be a much stronger force. For example, research suggests that under-nutrition during fetal life, such as during the Chinese famine of 1954 to 1964, raises the risk of diabetes in adulthood, especially when individuals live in nutritionally "rich" environments later in life (Berna et al., 2014).

2.10. Applications Covered of Factor Analysis on Medical Data

Application 1

Problem Statement: A confirmatory factor analysis of the metabolic syndrome in adolescents: an examination of sex and racial/ethnic differences.

According to (Gurka et al., 2012), the diagnosis of (MS) is typically based on cut-off points for various components, e.g. waist circumference and blood pressure. Because current MS criteria result in racial/ethnic discrepancies, their goal was to use confirmatory factor analysis to delineate differential contributions to MS by sub-group.

Research Design and Methods

Using 1999–2010 data from the National Health and Nutrition Examination Survey (NHANES), they performed a confirmatory factor analysis of a single MS factor that allowed differential loadings across sex and race/ethnicity, resulting in a continuous MS risk score that was sex and race/ethnicity-specific.

Results

Loadings to the MS score differed by racial/ethnic and gender subgroup with respect to triglycerides and HDL-cholesterol. ROC-curve analysis revealed high area-under-the-curve concordance with MS by traditional criteria (0.96), and with elevations in MS-associated risk markers, including high-sensitivity C-reactive protein (0.71), uric acid (0.75) and fasting insulin (0.82). Using a cut off for this score derived from ROC-curve analysis, the MS risk score exhibited increased sensitivity for predicting elevations in ≥2 of these risk markers as compared with traditional pediatric MS criteria.

Conclusions

The equations from this sex- and race/ethnicity-specific analysis provide a clinically-accessible and interpretable continuous measure of MS that can be used to identify children at higher risk for developing adult diseases related to MS, who could then be targeted for intervention. These equations also provide a powerful new outcome for use in childhood obesity and MS research by Gurka et al. (2012).

Application 2

Problem Statement: Confirmatory Factor Analysis used to determine the association of Adipocytokines with the Metabolic Syndrome.

Smits et al., (2013) used Confirmatory Factor Analysis (CFA) to test the hypothesis of whether adipocytokines were associated with the risk factor cluster that characterizes the metabolic syndrome (MS) or not. During the exercise, data from 134 nondiabetic subjects were analyzed using CFA. Insulin sensitivity (S_I) was quantified using intravenous glucose tolerance tests, visceral fat area by CT scan and fasting HDL, triglycerides, monocyte chemo-attractant protein-1 (MCP-1), serum amyloid A (SAA), tumor necrosis factor-α (TNF-α), adiponectin, resistin, leptin, interleukin-6 (IL-6), C-reactive protein (CRP) and plasminogen activator inhibitor-1 (PAI-1) were measured.

Results

The basic model representing the MS included six indicators comprising obesity, S_I, lipids and hypertension, and demonstrated excellent goodness-of-fit. Using multivariate analysis, MCP-1, SAA and TNF-α were not independently associated with any of the MS variables. Adiponectin, resistin, leptin, CRP and IL-6 were associated with at least one of the risk factors, but when added to the basic model, it decreased all goodness-of-fit parameters. PAI-1 was associated with all cardio-metabolic factors and improved goodness-of-fit compared to the basic model.

The conclusion was that the addition of PAI-1 increased the CFA model goodness-of-fit compared to the basic model, suggesting that this protein may represent an added feature of the MS. This was a demonstration of the power of Confirmatory Factor Analysis in determining the association between variables even in tricky situations.

Application 3

Should BMI or Waist Circumference Cut Points Be Ethnicity Specific?

These findings have sparked off an international debate about whether the cut points for overweight and obesity should be lower for Asians than for other ethnic groups. In 2004, the World Health Organization weighed the evidence on Asians' higher risk of weight-related diseases at lower BMIs. It declined to set different cutoff points for Asians, citing a lack of agreement among researchers as to what those lowered cutoffs should be. With the emergence of more research, however, several groups have begun to set lower cutoff points for BMI and abdominal obesity metrics among Asians. China and Japan define overweight as a BMI of 24 or higher and obesity a BMI of 28 or higher; in India, overweight is defined as a BMI of 23 or higher, and obesity, a BMI of 27 or higher. And the International Diabetes Federation now includes ethnic-specific criteria for the definition of abdominal obesity (Misra and Shrivastava, 2013) and (Prasad et al., 2011).

Application 4

Patients at risk of CVD and T2D

The concept of metabolic syndrome has several practical uses. One important use is in the everyday clinical assessment of patients, to identify patients at higher risk of T2D or CVD. However, the metabolic syndrome should not be considered only as a way to identify patients at increased risk, as other established risk assessment methods take other important factors into consideration (Meigs, 2004). For example, none of the definitions of metabolic syndrome take into account family history of diabetes, which is one of the most potent known T2D risk factors.

Thus, determination of metabolic syndrome would be inferior to the use of a specific risk assessment method such as the diabetes predicting model, which takes family history into account. Similarly, the metabolic syndrome definitions do not consider age, gender (although some of the cut points are gender specific), smoking, low-density lipoprotein (LDL) or total cholesterol levels, all known to be important CVD risk factors. Thus, metabolic syndrome would be inferior to a risk assessment tool, such as the Framingham risk score, for the prediction of CVD risk. The major use of metabolic syndrome is not so much in

identifying patients at general risk of CVD and T2D, but that it identifies a specific subgroup of patients with a shared pathophysiology. Thus, the term serves as shorthand for clinicians for the common underlying biological processes (Liu et al., 2015).

The NCEP ATP III definition is applied easily in the clinical setting. Physicians can easily score patients (and, indeed, motivated patients can score themselves) on the five criteria using easily measured end points and come up with a 'yes' or 'no' answer as to whether metabolic syndrome is present. This differs from some of the more complicated risk calculation methods, which may require complicated algorithms or computation to come up with an answer. Although it has not been proven, the hope is that realization of a diagnosis of metabolic syndrome will motivate people and their physicians to take appropriate steps to reduce their risk of CVD and T2D. This may involve lifestyle modifications such as improved food choices and increased physical activities, and appropriate pharmacological management for the component criteria (Lloyd-Jones et al., 2010).

Application 5

Factor Structure Underlying Components of Allostatic Load

It was stated by (McCaffery et al., 2012), that Allostatic load is a commonly used metric of health risk based on the hypothesis that recurrent exposure to environmental demands (e.g., stress) engenders a progressive dysregulation of multiple physiological systems. Prominent indicators of response to environmental challenges, such as stress-related hormones, sympatho-vagal balance, or inflammatory cytokines, comprise primary allostatic mediators. Secondary mediators reflect ensuing biological alterations that accumulate over time and confer risk for clinical disease but overlap substantially with a second metric of health risk, the metabolic syndrome. Whether allostatic load mediators co-vary and thus warrant treatment as a unitary construct remains to be established and, in particular, the relation of allostatic load parameters to the metabolic syndrome requires elucidation.

Here, confirmatory factor analysis is used to test whether a single common factor underlies variation in physiological systems associated with allostatic load and whether allostatic load parameters continue to load on a single common factor if a second factor representing the metabolic syndrome is also modeled. Participants were 645 adults from Allegheny County,

PA (30–54 years old, 82% non-Hispanic white, 52% female) who were free of confounding medications.

Model fitting supported a single, second-order factor underlying variance in the allostatic load components available in this study (metabolic, inflammatory and vagal measures). Further, this common factor reflecting covariation among allostatic load components persisted when a latent factor representing metabolic syndrome facets was conjointly modeled. Overall, this study provides novel evidence that the modeled allostatic load components do share common variance as hypothesized. Moreover, the common variance suggests the existence of statistical coherence above and beyond that attributable to the metabolic syndrome.

Data Analysis

Confirmatory factor analysis was conducted based on Bentler and Weeks' model using the EQS program. Tests of significance were set at 0.05 (two-tailed). The ratio of cases to variables was over 50:1, and the ratio of cases to parameters was 16:1. Both were sufficient for conducting CFA. A chi-square test was used to evaluate the congruency between the hypothesized model and empirical data, although it is well recognized that chi-square tests are sensitive to large sample size.

As such, 3 other model fit indices were used: comparative fit index (CFI; 0.95 or above; indicative of good fit), average absolute standardized residuals (0.05 or less; indicative of good fit), and root mean square error of approximation (RMSEA; 0.05 or less; indicative of good fit) Age, sex and race were statistically adjusted in each analysis. As models including paced or un-paced respiration produced similar results and 20 participants were missing data for un-paced respiration, models with paced respiration were presented in the manuscript. A model substituting un-paced respiration was presented. The first step was a confirmatory factor analysis of the metabolic syndrome.

Sample Characteristics

Demographic characteristics of the sample and descriptive statistics of the metabolic, vagal and inflammatory variables were determined. Participants were on average 45 years of age,

82% non-Hispanic white, 52% female and 16% current smokers. The sample was on average in the overweight range with average serum lipids and blood pressures in the normal range.

2.11. Metabolic Syndrome and Environmental Factors

In summary, the central features of the metabolic syndrome are insulin resistance, visceral adiposity, atherogenic dyslipidemia and endothelial dysfunction. These conditions are interrelated and share common mediators, pathways and pathophysiological mechanisms. A comprehensive definition of the metabolic syndrome, expressed as simply as possible, would contain only these features. The requirement of multiple criteria would ensure the exclusion of people with individual components (e.g. isolated hypertension or isolated hyperlipidemia), as opposed to the composite pathophysiology discussed above. Inclusion of both TG and HDL criteria increases the specificity for atherogenic dyslipidemia, and inclusion of the blood pressure criterion ensures that the physiologic derangements are severe enough to have resulted in endothelial dysfunction.

Of the various definitions for metabolic syndrome, the NCEP ATP III definition is the easiest to apply clinically and epidemiologically, because it uses straightforward criteria that are measured readily. Despite the ongoing controversy about whether the concept of metabolic syndrome is useful, it clearly defines specific pathophysiological mechanisms that link the central features. Consideration of metabolic syndrome as a specific entity allows for research on the genetic basis for susceptibility to the syndrome, a better understanding of its underlying pathophysiology and the development of treatment approaches.

2.12. Different Approaches to Unravel the Latent Structure within Metabolic Syndrome

Two advanced techniques of clustering using VARCLUS and matroid methods were discussed and implemented on a metabolic syndrome data set to analyze the structure of ten metabolic risk factors by (Woolston et al., 2012). Data were collected on subjects who were selected from the normative aging study based in Boston, Massachusetts. The sample included a total of 847 men aged between 21 and 81 years who provided complete data on selected risk factors during the period 1987 to 1991.

The researchers had noted that Exploratory Factor Analysis was understood to be a commonly used statistical technique in metabolic syndrome research to uncover latent structure amongst metabolic variables. The application of factor analysis requires methodological decisions that reflect the hypothesis of the metabolic syndrome construct. These decisions often raise the complexity of the interpretation from the output. They proposed two alternative techniques developed from cluster analysis which could achieve a clinically relevant structure, whilst maintaining intuitive advantages of clustering methodology.

According to their final results, four core components were identified by the clustering methods. They were labelled; obesity, lipids, insulin resistance and blood pressure. The exploratory factor analysis with oblique rotation suggested an overlap of the loadings identified on the insulin resistance and obesity factors. The VARCLUS and matroid analyses separated these components and were able to demonstrate associations between individual risk factors. An oblique rotation can be selected to reflect the clinical concept of a single underlying syndrome, however the results were often difficult to interpret. Factor loadings must be considered along with correlations between the factors, they reckoned. The correlated components produced by the VARCLUS and matroid analyses were not overlapped, which allowed for a simpler application of the methodologies and interpretation of the results. These techniques encouraged consistency in the interpretation whilst remaining faithful to the construct under study (Woolston et al., 2012).

2.13. Management of Metabolic Syndrome

The successful management of metabolic syndrome basically hinges on lifestyle, modification and pharmacological intervention. While attempts are ongoing in search for an approach that can simultaneously affect all the components, the current approach remains to treat each component as it becomes manifest. Since it is believed that a major driving force in Africa is epidemiologic transition, reverting back to African traditional lifestyles is a potential point of action to prevent the development of metabolic syndrome in Africans. This will involve paying attention to our local cardio-protective diets once again and improving on our level of physical activity structured into the activities of daily living of the individuals. This, however, calls for plans to increase awareness among African subjects and for those

already on treatment to adhere strictly to their medications. These actions constitute significant primary and secondary prevention strategies (Okafor, 2012).

Pharmacologically, several classes of drugs which include antihypertensive agents, oral glucose lowering agents, insulin sensitizers, and lipid-lowering agents are available to treat metabolic syndrome. Due to the clustering of the components of the syndrome, an individual with full-blown syndrome is exposed to a high pill burden and in turn, an increased cost. These can contribute to poor adherence or compliance. The earlier the components are addressed the better, however prevention still remains the watchword.

2.14. Exploratory and Confirmatory Factor Analysis of the PROMIS Pain Quality Item Bank

As part of the PROMIS project, (Revicki, Cook and Amtmann, 2014) developed a pool of 37 pain quality items, based on a review of existing pain questionnaires and development of new items. The assessment of pain sensation and quality was a key component in understanding the experience of individuals with chronic pain. The study evaluated the factor structure of the Patient-reported Outcome Measurement Information System (PROMIS) pain quality item bank. Methods and materials adopted included the utilization of a Web-based survey which was designed and completed by 845 members of the general population and 967 individuals with different types of chronic pain. Exploratory factor analysis (EFA) was conducted on a random split-half sample of the data to examine the factor structure of the 37 PROMIS pain quality items in the general population and in a chronic pain sample. A confirmatory factor analysis was also conducted in the holdout sample.

There is heightened recognition that pain is a multidimensional experience which has increased interest in measuring distinct aspects of the pain experience. Careful, self-report-based assessments of pain qualities are important for several reasons. First, such assessments may help clinicians better characterize and differentiate the unique qualities of pain associated with specific pain syndromes. Second, the quality of a patient's pain may provide clues to its underlying etiology. Finally, there is evidence that certain interventions

relieve pain because they alter the affective quality of pain, while other interventions relieve pain because they alter the sensory intensity of pain.

The EFA of the pain quality items resulted in comparable six-factor solutions for the general and chronic pain samples: (1) pulling/tugging pain; (2) tingling/numbness pain; (3) sharp/stabbing pain; (4) dull/aching pain; (5) pounding/pulsing pain; and (6) affective pain. The confirmatory factor analysis in the holdout sample supported this factor structure. The group of researchers recommended further research to evaluate the psychometric characteristics of the derived scales based on their factor scores.

This researcher, once again, was of the view that at this point in time, not enough emphasis was given to the complete analysis of the pain data. This criticism came in the wake of serious underutilization of Confirmatory Factor Analysis. A part from confirmation of the findings at the Exploratory Factor Analysis level, a further consideration should have been to extent the analysis and include a path analysis and establish the relationship between factors determined at the EFA level and their respective indicator variables. A diagram showing loadings between indicator variables and factors would have sufficed.

2.15. Association between Total and Abdominal Adiposity and Inflammation

Existence of Association between Total and Abdominal Adiposity and Inflammation in Older Adults Using a Factor Analysis Approach

According to (Brinkley et al., 2012), Obesity-related increases in multiple inflammatory markers may contribute to the persistent subclinical inflammation common with advancing age. However, it was unclear if a specific combination of markers reflected the underlying inflammatory state. These researchers used factor analysis to identify inflammatory factor(s) and examine their associations with adiposity in older adults at risk of disability. For methods and materials, they adopted; Adiponectin, CRP, IL-1ra, IL-1sRII, IL-2sRα, IL-6, IL-6sR, IL-8, IL-15, sTNFRI, sTNFRII, and TNF-α were measured in 179 participants from the Lifestyle Interventions and Independence for Elders Pilot (Mean ± *SD* age 77 ± 4 years, 76% white, 70% women). Body mass index, waist circumference, and total fat mass were assessed by anthropometry and dual-energy x-ray absorptiometry.

Results showed that IL-2sRα, sTNFRI, and sTNFRII loaded highest on the first factor (factor 1). CRP, IL-1ra, and IL-6 loaded highest on the second factor (factor 2). Factor 2, but not factor 1, was positively associated with 1-*SD* increments in waist circumference (β = 0.160 ± 0.057, p = .005), body mass index (β = 0.132 ± 0.053, p = .01), and total fat mass (β = 0.126 ± 0.053, p = .02) after adjusting for age, gender, race/ethnicity, site, smoking, anti-inflammatory medications, comorbidity index, health-related quality of life, and physical function.

These associations remained significant even after further adjustment for grip strength, but only waist circumference remained associated with inflammation after adjusting for total lean mass. There were no significant interactions between adiposity and muscle mass or strength for either factor. Several mechanisms for this age-related subclinical inflammation have been postulated, including increases in adipose tissue mass. Adipose tissue expresses and releases a number of inflammatory cytokines in direct proportion to the amount of adipose mass.

In addition, circulating levels of inflammatory markers were understood to be elevated in obesity and correlate with Body Mass Index (BMI), total body fat, and abdominal fat. Given that cytokines may accelerate adverse changes in body composition that are typical of the aging process, older adults may be vulnerable to obesity-related increases in inflammation. In this regard, aging is associated with a geriatric syndrome characterized by the coexistence of obesity and muscle impairment, either defined by poor muscle strength or low muscle mass. Given the results, it was concluded that greater total and abdominal adiposity were associated with higher levels of an inflammatory factor related to CRP, IL-1ra, and IL-6 in older adults, which may provide a clinically useful measure of inflammation in the target population.

2.16. Factor Analysis on Medical Data

A number of different statistical procedural methods have been employed to clearly scrutinize and bring out the information which is concealed in a variety of variables observed/collected on many human subjects (Gurka et al. 2013), (Furguson et al., 2009), (Woolston et al., 2012), (Martinez-Vizcaino et al., 2010) and (Huo et al. 2013). A large

portion of these approaches are multivariate statistically oriented. When the number of variables is greater than two, employment of multivariate analysis techniques gives simpler and more easily interpretable results for the evaluation of the observed matrix of data. In this study, it was attempted to determine latent factors that could explain the variability in a large set of data collected on many individuals of mixed health statuses.

2.17. Metabolic Data Analysis and Factor Analysis

The metabolic syndrome is typically diagnosed based on abnormalities in specific clustered clinical measures that are associated with increased risk for coronary heart disease (CHD) and Type 2 diabetes mellitus (T2DM). However, current MS criteria result in racial/ethnic discrepancies. Confirmatory factor analysis (CFA) were used to delineate differential contributions to MetS by sub-group, and if contributions were discovered, develop sex and racial/ethnic-specific equations to calculate MS severity (Gurka et al. 2012).

It was further noted that the speech, spatial, and qualities of hearing questionnaire (SSQ) is a self-report test of auditory disability. The 49 items ask how well a listener would do in many complex listening situations illustrative of real life. The scores on the items are often combined into the three main sections or into 10 pragmatic subscales. A report of factor analysis of the SSQ was conducted to further investigate its statistical properties and to determine its structure (Akeroyd et al., 2014).

Design, Study Sample and Results

Statistical factor analysis of questionnaire data, using parallel analysis to determine the number of factors to retain, oblique rotation of factors, and a bootstrap method to estimate the confidence intervals. 1220 people attended MRC IHR over the last decade. Three clear factors were found, essentially corresponding to the three main sections of the SSQ. They were termed "speech understanding", "spatial perception", and "clarity, separation, and identification". Thirty-five of the SSQ questions were included in the three factors. There was partial evidence for a fourth factor, "effort and concentration", representing two more questions (Akeroyd et al., 2014).

Assessing Clustering of Metabolic Syndrome Components Available at Primary Care for Bantu Africans Using Factor Analysis in the General Population

A study was conducted by Nasila et al in 2013 on data received from Kinshasa Hinterland, the Democratic Republic of Congo. The aim of this study was to understand, among other issues, the level of effectiveness of the application of Factor Analysis on anthropometric data. The data were collected on Bantu African participants who had different cutoffs of the components of MS. This was a cross-sectional, comparative as well as a correlational study. The cardio-vascular risk factors were defined in all, in MS group according to IDF cutoff points, and according to gender.

Out of 977 participants, 17.7%, 11% and 7.7% had T2DM, MS and CDM respectively. The study revealed that there were gender effects on all the variables. In the presence of MS, three factors were extracted which explained 75.1% of the total variance explained. Among those in the absence of MS, two factors were extracted which explained 48.1% of the total variance explained. The conclusion of the analysis was that the MS pathogenesis was more glucose-centered than it was abdominal obesity-centered.

2.18. Conclusion

This chapter has concerned itself with the literature review of this study. This has included mainly other works that have been dome and made known about the Metabolic Syndrome. This review has described, summarized, evaluated and clarified this literature included in this study. Chapter 2 has entailed literature review of Factor analysis and later a short review on work done on Metabolic Syndrome in relation to factor analysis.

Chapter 3

3. Research Methodology

3.1. Introduction

This chapter provided information on the thesis progress and procedures to deal with the data collection, ethical issues and the techniques for analysis. The techniques were explained, details provided for examiners and other readers with better understanding of the philosophy and the directions of the expression of views on this thesis. In this chapter the research methodology used in the study is described. The geographical area where the study was conducted, the study design and the population and sample are described. The procedure of data collection and some of the data variables have been described.

This research was indented to provide a step-by-step description of the application of factor analysis and interpretations of the results based on the output of data collected on anthropometric parameters (waist circumference and body mass index), blood pressure and plasma glucose in the general population; for people of both gender (men and women), and for both rural and urban inhabitants with different types of Diabetes Mellitus. Apart from the variables stated here, several other variables were included in the data collection exercise.

For ease of scrutiny and for more clarity of the demonstration of the procedure of factor analysis in medical circles, it was found proper and equally convenient to include the above stated six variables. Apparently, the above variables were selected with the prior knowledge of their relatedness.

3.2. Materials and Method

The initial study was a cross-sectional survey conducted between January, and April 2005, in Kinshasa Hinterland, in the Democratic Republic of Congo. The survey was specifically and extensively designed using a statistical multistage and stratified random model at each level to recruit a study sample with similar and representative characteristics of Kinshasa Hinterland demographic and socioeconomic structures whose results are quite comparable with global data on DM. Metabolic Syndrome was defined by the IDF criterion. The data was drawn from four different regions using the same collection method.

Variables such as weight, height, waist circumference (WC), systolic blood pressure (SBP), diastolic blood pressure (DBP), plasma fasting glucose and plasma post load glucose were collected following medically approved procedures and such procedures as have been stated above. It is reasonable to mention here that some of the variables were calculated as functions of other variables. Thus the variables collected were grouped into two types; primary and secondary variables. Primary variables include; systolic blood pressure, diastolic blood pressure, glucose fasting, etc. Secondary variables included; Body mass index (BMI) which was obtained by dividing weight (kg) by height (m^2). Cardio-metabolic waist circumference risk was defined by two components of MS such as concurrent presence of high blood pressure (SBP>=130mmHg or DBP >=85mmHg) and DM.

The history of anthropometry includes and spans from various concepts, both scientific and pseudoscientific, such (WHO Expert Committee, 1995) as; craniometry, paleoanthropology, biological anthropology, phrenology, physiognomy, forensics, criminology, phylogeography, human origins, and cranio-facial description, as well as correlations between various anthropometrics and personal identity, mental typology, personality, cranial vault and brain size, and other factors.

Throughout various times in history, applications of anthropometry have stretched vastly; from accurate scientific description and epidemiological analysis to rationales for eugenics and overtly racist, social movements—and its points of concern have been numerous, diverse, and sometimes highly unexpected. Below, the researcher lists and explains different anthropometric factors and how they were measured for this research. Human height varies greatly between individuals and across populations for a variety of complex biological, genetic, environmental, and other factors. Due to methodological and practical problems, its measurement is also subject to considerable error in statistical sampling.

The average height in genetically and environmentally homogeneous populations is often proportional across a large number of individuals. Exceptional height variation (around 20% deviation from a population's average) within such a population is sometimes due to gigantism or dwarfism, which is caused by specific genes or endocrine abnormalities.

BMI was used as an indicator of both underweight, overweight and obesity. On a population basis, in adults, there is a strong association between BMI and health risk. There are however many individuals for whom BMI is an inappropriate measure of body fatness. These are individuals whose high body mass index is due to excess muscle rather than excess body fat or in those with osteoporosis who will have a lower than usual BMI, or those who have a different body build (individuals with unusually long or short legs or a different body fat distribution) (WHO Expert Committee, 1995).

3.3. Research Design

A cross-sectional survey collects data to make inferences about a population of interest (universe) at one point in time. Cross-sectional surveys have been described as snapshots of the populations about which they gather data. Cross-sectional surveys may be repeated periodically; however, in a repeated cross-sectional survey, respondents to the survey at one point in time are not intentionally sampled again, although a respondent to one administration of the survey could be randomly selected for a subsequent one. Cross-sectional surveys can thus be contrasted with panel surveys, for which the individual respondents are followed over time. Panel surveys usually are conducted to measure change in the population being studied. Cross-sectional surveys can be conducted using any mode of data collection, including telephone interviews in which landline telephones are called, telephone interviews in which cell phones are called, face-to-face interviews, mailed questionnaires, other self-administered questionnaires, electronic mail, web data collection.

This study was a cross-sectional survey conducted between January, and April 2005, in Kinshasa Hinterland with details previously published. The survey was specifically and extensively designed using a statistical multistage and stratified random model at each level to recruit a study sample with similar and representative characteristics of Kinshasa Hinterland demographic and socioeconomic structure and results comparable with global data on DM. Each region contributed a number of cluster (EDs) calculated by population number: 185, 112 inhabitants for the upper urban area of Gombe, 161,410 inhabitants of the semi-rural Kisero area, 153,265 inhabitants for the urban Lukemi area and 146,034 inhabitants for the deepest rural Feshi area. The sample size was calculated as $Z^2 * P * Q$

where P is the expected prevalence of DM in each area, $Q = 1 - P$, P is the absolute accuracy of 2% and f = 8.5 to correct the design effect.

The details of collection of weight, height, waist circumference (WC), systolic blood pressure (SBP), diastolic blood pressure (DBP), plasma fasting glucose and plasma post load glucose have been described elsewhere (Olinto et al., 2011). Body mass index (BMI) was obtained in dividing weight (kg) by height in m^2. In our setting with limited resources, this explains how the data collection procedure.

3.4. Data Collection

The survey was specifically and extensively designed using a statistical multistage and stratified random model at each level to recruit a study sample with similar and representative characteristics of Kinshasa Hinterland demographic and socioeconomic structures whose results are quite comparable with global data on DM. MS was defined by the IDF criterion.

The details of collection of weight, height, waist circumference (WC), systolic blood pressure (SBP), diastolic blood pressure (DBP), plasma fasting glucose and plasma post load glucose have been described elsewhere. BMI should be derived from the data entry of weight and height. It should be stored on the raw data set as a continuous variable and should not be aggregated or rounded. It is recommended that in population surveys, socio-demographic data including ethnicity should be collected. Date of birth and date of measurement should also be recorded. This allows decimal age to be calculated if required.

Anthropometry is a branch of anthropology concerned with comparative measurements of the human body and its parts as well as the variables which impact these measurements. Anthropometric data consists of collections of measurements, often presented in tabular format or annotated diagrams of human figures. The primary dimensions measured are of bones, muscle, and adipose tissue. This data is used in human factors/ergonomics applications in order to ensure that designs and standards are realistic.

Data Collection Procedure

This study was a cross-sectional survey conducted between January, and April 2005, in Kinshasa Hinterland. This study was carried out in compliance with the Helsinki Declaration (59th WMA General Assembly, Seoul, South Korea, October 2008. The data collection was approved by the Ethics Committee of Lomo Medical Clinic (Ref-00038-03-07) at Kinshasa Limetè. Fully informed and written consent was obtained from all participants who were adults. The survey was specifically and extensively designed using a statistical multistage and stratified random model at each level to recruit a study sample with similar and representative characteristics of Kinshasa Hinterland's demographic and socioeconomic structure. It is the researcher's belief that any analysis results would be comparable with other Diabetes Mellitus data at global level. The target area was sub-divided into different strata. Each region contributed on the basis of its population size. The target area was populated by: 185, 112 inhabitants occupied the upper urban area of Gombe, 161,410 inhabitants originated from the semi-rural Kisero area, 153,265 inhabitants for the urban Lukemi area and 146,034 inhabitants for the deepest rural Feshi area. The variables collected included: weight, height, waist circumference (WC), systolic blood pressure (SBP), diastolic blood pressure (DBP), plasma fasting glucose and plasma post load glucose. Derived variables such as body mass index (BMI) was obtained in dividing weigh (kg) by height (m^2). In the data collection setting with limited resources and lack of routinely measured insulin resistance (gold standard), the criteria of MS diagnosis proposed by the International Diabetes Federation (IDF) was followed as follows: raised systolic blood pressure (SBP > 130 mmHg) and diastolic blood pressure (DBP > 85 mmHg), elevated triglycerides (TG > 1.7 mmol/L), low high-density lipoprotein cholesterol (HDL < 1.04 mmol/L in men and <1.29 mmol/L in women) levels, abdominal obesity defined by increased waist circumference (WC > 94 cm in men and >80 cm in women), and fasting plasma glucose (FPG > 5.6 mmol/L)(6). CDM was defined by the constellation of 3 components of WHO–defined MS such as diabetes, hypertension and BMI > =30 kg/m2. However, the absence of CDM was defined among participants without pre-hypertension, abdominal obesity, BMI > =25 kg/m2, and CDM. The definition of diabetes was based on clinical arguments and the latest WHO/IDF criteria among persons with the fasting venous plasma glucose level > =126 mg/dL or Post-load venous blood plasma level > =200 mg/dL.

This was an undiagnosed T2DM so that information about HbA1c, duration of diabetes, and medications was not available and compulsory.

3.5. Theoretical Fundamentals of Factor Analysis

This subsection presents recent work on the factor analysis that will be applied on Medical Syndrome data that was obtained from Kinshaha, DR of Congo. This part of the thesis aims at discussing recent work done on the factor analysis methodology, specifically focusing on Explanatory and Confirmatory Factor Analysis. The methodology discussed in this section is an extract taken from Kline (2013) and Kline (2010, ch 3) and other recent statistical articles and journals such as (Williams, Onsman and Brown, 2010).

3.5.1. Principles and Explanations of Factor Analysis

While factor analysis has origins dating back 100 years through the work of Pearson and Spearman, the practical application of this approach has been suggested to be in fact a modern occurrence. Spearman (1904a) is also credited with articulating basic principles of classical measurement theory and, not surprisingly, there is close connection between factor analysis and psychometrics. In the latter observation were made and it was understood that an observed score x_{ij} for item i measured at time j is understood to be composed of two components T_i and a residual component E_{ij}, which can be statistically expressed as follows;

$$X_{ij} = T_i + E_{ij} \tag{1}$$

Since measurement error is random and thus unrelated to true scores, variance in observed scores can be broken down into two non-overlapping parts, which are

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \tag{2}$$

Score reliability coefficients calculated in samples estimate the ratio of true score variance over total observed variance as follows;

$$r_{XX} = \hat{\sigma}_T^2 / \hat{\sigma}_X^2 \tag{3}$$

And the expression $1 - r_{XX}$ estimates the proportion of total variance due to measurement error. For examples if $r_{XX} = 0.65$, then $1 - r_{XX} = 0.35$ or 35% of total variance is caused by the residual.

Factor analysis further partitions true variance into *common variance* and *specific variance*. Common variance is shared among a set of indicators and is a basis for inter-correlations among them that depart appreciably from zero. In factor analysis, it is generally assumed that (a) common variance is due to the effects of underlying factors and (b) the number of factors of substantive interest is less than the number of indicators. It is impossible to estimate more common factors than indicators, but for parsimony's sake, there is no point in retaining a model with just as many explanatory entities (factors) as there are entities to be explained (indicators) (Mulaik, 2009). The goal of most factor analyses is thus to identify and interpret a smaller number of factors that explains most of the common variance.

The statistics h^2 estimates the proportion of total variance that is common and it is referred as to communality. For example, if h^2 = .60, then 60% of total indicator variance is common and thus potentially explained by underlying factors. The rest, or 30% of the total variance, is unique variance, which is made up of specific variance (systematic but unshared) and measurement (random) error. Specific variance is not explained by common factors; instead, it may be due to characteristics of individual indicators, such as the particular stimuli that make up a task, that also affect observed scores. The various partitions of standardized total indicator variance in factor analysis just described is illustrated in Figure 2 below. Note in the figure that as the proportion of error variance increases, the proportion of systematic (true) variance decreases, which can in turn reduce the overall proportion of common variance.

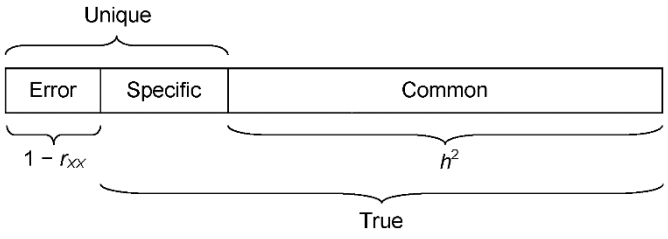

Figure 2 Basic partition of standardized indicator variance in factor analysis. h^2, proportion of Common variance, or communality; r_{XX} , score reliability coefficient (Kline, 2010).

Statistical analysis of unreliable scores usually leads to inaccurate results, and factor analysis is no exception. In general, score reliabilities that exceed .90 are considered excellent, coefficients in the range of about .80 are considered good, but coefficients appreciably less than .70 are potentially problematic. If $r_{XX} < .50$, then most of total variance is due to measurement error. Indicators with such low score reliabilities should be excluded from the analysis.

3.5.2. Fundamentals of Factor Analysis

We need more than just explained in introductory topics. The argument is that the really relationships between these variables are driven by underlying latent variables which measure the degree of a set of variables in each factor. The measurement is so designed as to construct different latent variables for different groups of variables (Mulaik, 1987). This leads us to factor analysis. Consider turning the variables x_1 and x_2 into standardized variables, z_1 and z_2. Assume then, that the actual observed values of these variables are determined by some common underlying factor (F) and a unique factor for each variable, y_1 and y_2. Then we have;

$$z_{1i} = b_1 F_i + u_1 y_{1i} \tag{4}$$

$$z_{2i} = b_2 F_i + u_2 y_{2i}$$

We will assume that F and both $y\,'\,s$ are standardized and that they are all uncorrelated with one another.

Let's ask what the variance of z_1 is? Since, z_1 is in standard form, this variance is simply given by $E(z_1^2)$, so:

$$E(z_1^2) = E(b_1^2 F^2 + u_1^2 y_1^2 + 2b_1^2 u_1^2 F y_1) \tag{5}$$

$$E(z_1^2) = E(b_1^2 F^2 + u_1^2 y_1^2 + 2b_1^2 u_1^2 F y_1)$$

$$E(z_1^2) = b_1^2 E(F^2) + u_1^2 E(y_1^2) + 2b_1^2 u_1^2 E(F y_1)$$

$$var(z_1) = b_1^2 var(F) + u_1^2 var(y_1) + 2b_1^2 u_1^2 covar(F, y_1)$$

Assuming variances are one and covariance is zero (assumed variable independence):

$$var(z_1) = b_1^2 + u_1^2 \tag{6}$$

$$1 = b_1^2 + u_1^2$$

Therefore, the variance in variable z_1 is determined by the contribution of the common factor and the unique factor. What about the covariance between z_1 and F:

$$covar(z_1, F) = E(z_1 F) \tag{7}$$

$$covar(z_1, F) = E((b_1 F + u_1 y_1)F)$$

$$covar(z_1, F) = (b_1 E(F^2) + u_1 E(y_1)F)$$

$$covar(z_1, F) = (b_1 var(F) + u_1 covar(y_1), F)$$

$$covar(z_1, F) = b_1$$

Since both F and z_1 are standardized, the covariance is the same as the correlation. So the correlation coefficient r between z_1 and F is simply given by b_1.

How about the covariance between z_1 and z_2 is;

$$covar(z_1, z_2) = E(z_1 z_2) \tag{8}$$

$$covar(z_1, z_2) = E((b_1 F + u_1 y_1)(b_2 F + u_2 y_2))$$

$$covar(z_1, z_2) = E(b_1 b_2 F^2 + b_1 u_1 F y_1 + u_1 u_2 y_1 y_2)$$

$$covar(z_1, z_2) = b_1 b_2 var(F) + b_1 u_1 cov(F, y_1) + u_1 u_2 cov(y_1, y_2)$$

$$covar(z_1, z_2) = b_1 b_2$$

$$r_{z_1, z_2} = b_1 b_2$$

So the correlation matrix between a set of variables is completely determined by their common factors.

Given the decomposition of the variance in z_1, we can define the communality (h_2^j) of each variable as b_j^2. This is the proportion of the variance explained by the common factor. The uniqueness of the variable is given by $1 - h_2^j$.

This framework can be generalized to a set of j variables and m factors.

$$z_{ji} = b_{j1} F_{1i} + b_{j2} F_{2i} + \cdots + b_{jm} F_{mi} + u_j Y_{ji} \quad (9)$$

With multiple common factors, communality is given by:

$$h_j^2 = \sum_{p=1}^{m} b_{jp}^2 \quad (10)$$

This setup is called common-factor analysis (Fan et al., 2013), (Velicer & Jackson, 2009), (Russell, 2002) and (MacCallum, 1983). One can also assume that factors explain everything and there are no unique factors. In this case the number of factors will equal the number of variables. The difference between the methods is that the common-factor approach only attempts to explain the covariation between observed variables, the principal-component approach attempts to explain all of the variation in the observed variables (McArdle et al., 2014) and (Velicer and Jackson, 2009).

We refer to the $b's$ as factor loadings. They tell us the correlation coefficient between each factor and the observed variables. The values of each observation on the factor F_k are referred to as factor scores. The fundamental problem is that we never have the factors, we have the observed correlation matrix between variables and we assume it may be produced

by a set of common factors. If we are given a set of factors and their factor loadings for certain variables, then we can reproduce the one and only one correlation matrix for these variables, but given a correlation matrix, we cannot deduce a single set of factor loadings (McArdle et al., 2014) and (Velicer and Jackson, 2009).

3.5.3. Decision Points in Factor Analysis

Factor analysis consist three basic analytical decisions which can be summarized as are as follows: (Kline, 2010), (Kline, 2012) and (Kline, 2013),

- Whether factor analysis is the appropriate technique and, if so, whether to use EFA or CFA
- The set of indicators to be analyzed and the composition and size (N) of the sample
- The data matrix to be analyzed; the basic choice is between a correlation matrix versus a covariance matrix

When Factor Analysis is Appropriate

The decision whether to conduct a factor analysis is usually not complicated. This is because the technique's basic purpose description of latent variables that explain observed correlations is pretty straightforward. Differences between EFA and CFA were considered in the previous section. The technique of EFA may be a better choice in less mature research areas where basic measurement questions are not yet resolved. It also requires fewer a priori assumptions than CFA, which tests stronger hypotheses than EFA. In assessment research, EFA tends to be used in earlier studies and CFA in later studies in the same area. The selection of EFA or CFA implies additional decision point specific to each technique that are explained later in this chapter. Considered next are decisions that apply to both EFA and CFA, refer (Kline, 2010) and (Kline, 2013).

Indicator Selection

The selection of indicators is critical because the quality of what comes out of a factor analysis (the results) depends heavily on the nature and quality of scores analyzed. Summarized next are recommendations by Fabrigar et al. (1999) concerning indicator selection: First, the researcher must define the hypothetical constructs of interest. For example, if the goal is to delineate dimensions of anxiety, then the researcher should consult relevant theoretical and empirical works about the nature and number of factors, such as state anxiety, trait anxiety, anticipatory anxiety, social anxiety, and so on. Next, candidate indicators that as a set adequately sample the different domains should be identified. Ideally, not all indicators will rely on the same method of measurement, such as assessment of anxiety through self-report questionnaires only. This is because common method variance can affect all scores regardless of common latent variables. For instance, it is frequent in anxiety studies to measure physiological variables, such as heart rate or galvanic skin response, in addition to self-report, refer (Kline, 2010) and (Kline, 2013).

It is also generally necessary to select multiple indicators of each presumed dimension. Multiple-indicator measurement not only tends to sample more facets of constructs of interest, but technical problems in the analysis are more likely to happen if some factors have too few indicators. This is especially true in small samples where some factors have just 2 indicators. In general, there should at least 3–5 indicators for each anticipated factor. If a total of four anxiety dimensions are expected, for instance, then the minimum number of indicators would about 12–20. But sometimes there are few theoretical or empirical bases for predicting the number of factors before conducting the analysis. In this case, the researcher must nevertheless try to delineate the population of indicators and then measure as many as possible in a sufficiently large sample (Fabrigar et al., 1999). It is also crucial to select indicators with good psychometric characteristics.

As in most behavioral science studies, the sample should be representative of the population to which the results should generalize. For factor analysis, the sample should also be (a) relatively heterogeneous on the indicators but (b) relatively homogenous on other variables that do not co-vary substantially with the indicators (Mulaik, 2009). Because factor analysis is essentially a regression technique where the predictors are latent variables, its results can

be distorted by range restriction. Suppose that a researcher administers cognitive and scholastic ability tests within a sample of school children enrolled in programs for the gifted. Because the range of individual differences among gifted children on these may be relatively narrow, absolute magnitudes of inter-correlations among the tests may be restricted compared with a general sample of students. Because correlation is the "fuel" of factor analysis, results in range-restricted samples may not be very meaningful. However, relative homogeneity among participants on other variables, such as demographic characteristics, that are not strongly related to the indicators helps to ensure that the factors affect scores of all cases the same way, refer (Kline, 2010), (Kline,2012) and (Kline, 2013). That is, the same basic measurement model should hold for all cases (Mulaik, 2009).

Sample Size

A critical question concerns minimum sample sizes required for the analysis. In general, factor analysis is a large sample technique, so the more cases the better. (This assumes that a larger sample is just as representative as a smaller one.) Early sample size recommendations for EFA were based on ratios of the number of cases to the number of indicators. For example, the recommendation for a 10:1 ratio means that there are at least 10 cases for every indicator, so an analysis of 10 indicators would require a minimum sample size of100; a more stringent 20:1 ratio would require at least N = 200 for 10 indicators, and so on. There are two problems with such guidelines. First, there is no clear consensus in the literature about the optimal cases-to-indicators ratios for EFA. A 10:1 ratio is probably the most common guideline, but some methodologists advocate even higher ratios, such as 20:1. Second, sample size requirements depend on the population (true) factor model. Specifically, fewer cases are needed when each factor has at least 3–4 indicators and average communalities across the indicators are about .70 or higher (e.g., MacCallum, Widaman, Zhang, & Hong, 1999). In this ideal scenario, a 10:1 cases-to-indicators ratio may suffice, but absolute sample sizes less than 100 may be untenable in factor analysis. A minimum sample size of 200 seems more defensible. However, cases-to-indicators ratios that exceed 20:1 and minimum sample sizes of 400 or more may be required when the ideal conditions just listed do not hold (Fabrigar et al., 1999).

Results of some reviews suggest that sample sizes in published EFA studies are typically too small. For example, Costello and Osborne (2005) surveyed a total of 305 factor analytic studies published over a two-year period and listed in the PsychINFO database. Most of these analyses (63%) were conducted with cases-to-indicators ratios <10:1, and a total of 41% were based on ratios < 5:1. Only 21% of the studies featured cases-to-indicators ratios > 20:1. In a separate computer simulation study where factor analyses were conducted in generated samples of different sizes, Costello and Osborne (2005) found that most factor solutions based on cases-to-indicators ratios <10:1 were incorrect. When the ratio is 2:1, however, the rate of incorrect results was 90%, and almost one-third of these analyses failed due to technical problems.

Ratio-type recommendations for minimum sample sizes in CFA are not based on the number of indicators but instead on the number of parameters in the entire measurement model. In CFA, parameters include pattern coefficients, error variances and covariances (i.e., for correlated errors), and factor variances and covariances. Models with more parameters—even for the same number of indicators—require more estimates, so larger samples are necessary in order for the results to be reasonably precise. Sample size requirements in CFA also vary with the type of estimation method used and the distributional characteristics of the data. In general, somewhat smaller sample sizes are needed when the standard estimation method in SEM, maximum likelihood (ML) estimation, is used and the distributions are multivariate normal. In this case, a 20:1 ratio is recommended, that is, there should be at least 20 cases for each model parameter estimated in the analysis (e.g., Jackson, 2003). A "typical" sample size in SEM is about 200 (e.g., Shah and Goldstein, 2006), which may be adequate for analyzing a CFA model with 10 or so parameters. However, much larger sample sizes may be needed when a method other than ML estimation is used or distributions are severely non-normal. Another framework for estimating minimum sample sizes in CFA involves estimation of the statistical power of tests about either individual parameters or about the fit of the whole model to the data. A variation is to specify a target level of power, such as .80, and then estimate the minimum sample size needed for that target see Mac-Callum, Browne, and Sugawara (1996) and Kline (2010, chapter 8) for more information.

Data Matrix Analyzed

Most researchers input raw data files for computer statistical analyses. These same researchers may be surprised to learn that the raw data themselves are not necessary for most types of factor analysis. Specifically, if a raw data file is submitted, the computer will create its own matrix summary of the data, which is then analyzed. It is also possible in many computer tools to input a matrix summary instead of raw data. The capability to analyze summary statistics also provides the basis for a secondary analysis in which data collected by others are reanalyzed but where the raw data are unavailable. Many journal articles about the results of factor analysis contain enough information, such as correlations and standard deviations, to create a matrix summary of the data, which can then be submitted to a computer program for analysis. Thus, readers of these works can, with no access to the raw data, replicate the original analyses or estimate alternative models not considered in the original work. This is why it is best practice for researchers to report sufficient summary statistics for a future secondary analysis.

There are two basic types of matrix summaries of raw data, a Pearson correlation (r) matrix and a covariance (cov) matrix. The default matrix analyzed in most EFA computer procedures is a correlation matrix. Pearson correlations measure the degree of linear association between two continuous variables. Specifically, r measures the degree to which the rank order of scores on one variable corresponds to the rank order on the other variable also taking account of the relative distances between the scores. The entries in the diagonal of a correlation matrix all equal 1.0, which are also the variances of all variables in a standardized metric.

The default data matrix in SEM computer programs is the covariance matrix. This is because the standard method in SEM, ML estimation, analyzes unstandardized variables. It is possible in SEM to fit a CFA model to a correlation matrix, but special methods are needed (Kline, 2010, chapter 7). The diagonal entries in a covariance matrix are the variances of the indicators in their original (unstandardized) metrics. The off-diagonal entries are the covariances, which for two continuous variables X and Y is where r is the Pearson correlation and SDX and SDY are their standard deviations. A covariance thus represents the strength of the association between X and Y and their variabilities, albeit with a single

number. Because the covariance is an unstandardized statistic, its value has no upper or lower bound. For example, covariances of, say, –1,003.26 or 13.58 are possible. The statistic cov encapsulates all the information measured by r plus the degree of "spreadoutedness" of the scores on both indicators (Thompson, 2004).

Because the information conveyed by cov and r is not the same, it can happen that the fit of a measurement model to a correlation matrix is not the same as the fit of the same model to a covariance matrix in the same sample. Also, the factor analysis of a covariance matrix generates two sets of estimates, an unstandardized solution and a standardized solution. Only the latter is calculated by the computer when a correlation matrix is analyzed. For all these reasons, think carefully about the choice of which type of data matrix to analyze and report that choice in written summaries of the results. Finally, the analysis of correlations or covariances assumes that the indicators are continuous variables. This is most likely to be true when each indicator is a scale that generates a total score over a set of items. However, individual items with Likert-type response formats (e.g., 0 = disagree, 1 = uncertain, 2 = agree) are not continuous variables. Instead, they are generally considered to be ordinal, and their distributions tend to be non-normal. Therefore, analyzing a Pearson correlation matrix (or the corresponding covariance matrix) when the indicators are items may not be appropriate. The analysis of items in factor analysis is considered later in this chapter; the discussion that follows assume the analysis of scales.

3.5.4. Choosing the Number of Factors

Several criteria have been proposed for choosing m, the number of factors. We consider four criteria, which are similar to those given for choosing the number of principal components to retain. The four criteria are stated as follows:

- Choose m equal to the number of factors necessary for the variance accounted for to achieve a predetermined percentage, say, 80%, of the total variance $tr(S)$ or $tr(R)$.
- Choose m equal to the number of eigenvalues greater than the average eigenvalue. For R the average is 1; for S it is $\sum_{j=1}^{p} \theta_j / p$

- Use the scree test based on a plot of the eigenvalues of S or R. If the graph drops sharply, followed by a straight line with much smaller slope, choose m equal to the number of eigenvalues before the straight line begins.
- Test the hypothesis that m is the correct number of factors, H_0: $\Sigma = \Lambda\Lambda' + \varphi$, where Λ is $p\ x\ m$.

The first method applies particularly to the principal component method. As understood from previous discussions, the proportion of total sample variance (variance accounted for) due to the jth factor from S is $\sum_{j=1}^{p} \hat{\lambda}_{ij}^2 / tr(S)$.

3.5.5. Global Model Fitting

There are two known types of methods that are used to perform indices in literature used to perform a global model fit amongst others. The two types of indices are;

- Comparative (Incremental) Fit Indices (CFI) (Normal Fit Index)

  Which performs a fit relative to a "null' model (of 0 covariance) and relative to this assumption your model fits good.

  CFI Formulation:

  Bases on the idea of NCP $(\chi^2 - df)$

$$CFI = 1 - \frac{max[(\chi_T^2 - df_T), 0]}{max[(\chi_T^2 - df_T), (\chi_N^2 - df_N), 0]} \tag{11}$$

  From 0 to 1: bigger is better, anything which is greater than 0.90 is acceptance" and anything more than 0.95 is good.

- Turcker-Lewis Index (Non-normal Fit Index)

$$TLI = \frac{\left(\frac{\chi_N^2}{df_N}\right) - \left(\frac{\chi_T^2}{df_T}\right)}{\left(\frac{\chi_N^2}{df_N}\right) - 1} \tag{12}$$

From less than 0 to greater than 1, bigger is better, any value greater than 0.95 is good.

3.5.6. Types of Factor Analysis

As discussed there are two major classes of factor analysis: Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA), (Williams, Onsman and Brown, 2010). Their differences are outlined below in detail.

- Unrestricted measurement models are estimated in EFA, but it is restricted measurement models that are analyzed in CFA. This means that the researcher must explicitly specify the indicator-factor correspondence in CFA, but there is no option to do so in EFA, (Williams, Onsman and Brown, 2010).
- Unrestricted measurement models in EFA are not identified, which means there is no unique set of statistical estimates for a particular model. This property concerns the rotation phase, which is part of most applications of EFA. In contrast, CFA models must be identified before they can be analyzed, which means that there is only one exclusive set of estimates. Accordingly, there is no rotation phase in CFA, see Henson and Roberts (2006).
- It is assumed in EFA that the specific variance of each indicator is not shared with that of any other indicator. In contrast, CFA permits, depending on the model, estimation of whether specific variance is shared between pairs of indicators.
- Output from CFA computer procedures contains the values of numerous fit statistics that assess the fit of the whole model to the data. In contrast, fit statistics are not generally available in standard methods of EFA (including principle components analysis and principle axis factoring, defined later) carried out by computer programs for general statistical analyses, such as SPSS (IBM, Corp, 2012) and SAS/STAT (SAS Institute, Inc., 2012), but some more specialized computer programs, such as Mplus (Muthén & Muthén, 1998–2012), may print certain types of fit statistics for particular EFA methods.
- Procedures for EFA are available in many computer tools for general statistical analyses, such as SPSS and SAS/STAT. In contrast, more specialized computer tools for structural equation modeling (SEM) are needed for CFA because the latter is the SEM technique for estimating restricted measurement models. Some widely used SEM computer tools include LISREL (Jöreskog & Sörbom, 2012) and Mplus (e.g., Kline, 2010, Ch 4).

Presented in there Figure 9 below there are two hypothetical measurement models for six indicators and two factors represented with symbols from SEM. These include squares or rectangles for observed variables, ellipses or circles for latent variables or error terms, lines with a single arrowhead (→) for presumed direct effects from causal variables to variables affected by them, two-headed curved arrows that exit and re-enter the same variable ( ) for variances of factors or error terms; and curved line with two arrowheads ( ) for covariance's (in the unstandardized solution) or correlations (in the standardized one) between either pairs of factors or pairs of error terms (Kline, 2010, chap. 5).

Depicted in Figure 3 (a) is an unrestricted two-factor model of the kind analyzed in EFA. Without specific instruction from the user to do otherwise, an EFA computer procedure could theoretically generate all possible unrestricted factor solutions, which equals the number of indicators. The most basic solution is a single-factor model, which reflects the assumption that all indicators depend on just one common factor. Next is a two-factor model, then a three-factor model, and so on up to the most complex model possible with just as many factors as indicators. In practice, EFA computer procedures rely on default.

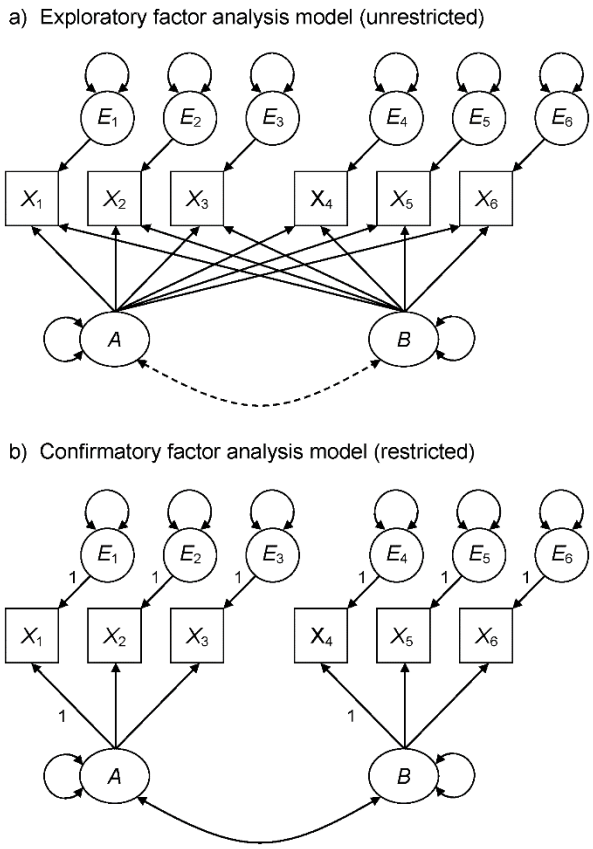

Figure 3 An Explanatory Factor Analysis and Confirmatory Factor Analysis for 6 indicators and 2 factors.

Statistical criteria for determining the number of factors to retain, but these defaults do not always correspond to best practice. These issues are elaborated later in the section about EFA, but the point now is that EFA does not require the researcher to specify the number of factors in advance.

The model in Figure 3 (a) is unrestricted concerning the correspondence between indicators and factors. In EFA, each indicator is regressed on every factor in the model. The statistical estimates for this part of EFA are actually regression coefficients that may be in either standardized or unstandardized form, just as in the technique of multiple regression. The difference is that predictors in ordinary regression are observed variables, but the predictors in EFA (and CFA, too) are latent variables. For example, there are two lines with single arrowheads that point to indicator X_1 from each of the factors in Figure 3 (a). These paths represent the presumed direct effects of both factors on X_1, and the proper name of the statistical estimates of these effects is referred to as factor loadings. The larger point is that all possible factor loadings are calculated for each indicator in EFA.

Pattern coefficients in factor analysis are interpreted just as coefficients in standard multiple regression. Suppose that the unstandardized pattern coefficient for the path $A \rightarrow X_1$ is 5.25 and that the standardized pattern coefficient for same path is .60. These results indicate that, respectively, (a) a difference of 1point on factor A predicts a difference of 5.25 points in the raw score metric of X_1, controlling for the other factors, and (b) a difference of a full standard deviation on factor A predicts a difference of .60 standard deviations on X_1, again controlling for all other factors. In multiple regression, standardized coefficients are not generally correlation coefficients when the predictors co-vary. It is only in the very special case where all inter-correlations among the predictors are zero that standardized regression coefficients are interpreted as correlations.

The same distinction holds in factor analysis: When the factors are correlated, the value of an indicator's standardized pattern coefficient will not typically equal that of its structure coefficient, which is the estimated correlation between an observed variable and a latent variable. It is only when the factors are independent that standardized pattern coefficients will equal the corresponding structure coefficient for every indicator. The failure to correctly distinguish standardized pattern coefficients from structure correlations when the factors are correlated can lead to misinterpretation of the results (Graham, Guthrie, & Thompson, 2003).

Each indicator in Figure 3 (a) has an error term, which represents unique variance in factor analysis (see Figure 2). The curved line with two arrowheads in the figure represents in EFA the possibility to estimate the correlation (or covariance in the unstandardized solution) between factors A and B. Because it is not required in EFA to estimate factor correlations, the curve of the symbol in Figure 3 (a) is presented as dashed instead of solid. When factor correlations are not estimated in EFA, it is assumed that the factors are orthogonal, or all pairwise independent. Most EFA computer procedures by default do not analyze correlated factors. Instead, the user must typically specify a rotation option that permits the factor to co-vary when the goal is to analyze correlated factors. Whether the computer default to analyze orthogonal factors makes theoretical sense in most applications of EFA is considered later.

Presented in Figure 3 (b) is a restricted measurement model for two factors and six indicators of the kind analyzed in CFA. In the guise of its covariance structure, the model in Figure 3 (b) represents the hypotheses that indicators $X_1 - X_3$ measure factor A, indica-tors $X_4 - X_6$ measure factor B, and the factors covary. This particular model is a standard CFA model with the characteristics listed next:

1. Each indicator is represented as having two causes; a single factor that the indicator is supposed to measure and all other unique sources of influence represented by the error term (e.g., $A \rightarrow X_1 \rightarrow E_1$). In CFA models with ≥ 2 factors, this property implies that some pattern coefficients are zero. For example, there is no path that connects factor B with indicator X_1 in Figure 3 (b). This specification reflects the hypothesis that X_1 does not depend on B. It also implies that the pattern coefficient for the direct effect of B on X_1 is assumed to equal zero and, consequently, the computer will not estimate this direct effect. As mentioned, it is the researcher who specifies the particular pattern coefficients to be estimated in CFA, not the computer as in EFA.
2. All possible pairs of factors are assumed to co-vary in CFA models. This assumption is represented in Figure 3 (b) by the specification A ↷ B, where the curve in the symbol ↷ is shown as solid instead of dashed. In models with ≥3 factors, every pair of factors is connected by the symbol ↷ for a covariance. Although it is theoretically possible in CFA to analyze a measurement models where some factors are assumed to be unrelated, such models are not standard models. In fact, some SEM computer tools automatically specify correlations between every pair factors in CFA measurement models.
3. The error terms in the CFA model of Figure 3 (b) are independent of each other, which reflects the same assumption as in EFA that indicator specific variances do not overlap. However, it may be possible in CFA to estimate error correlations (standardized) or error covariances (unstandardized) between some pairs of indicators. Suppose that X_1 and X_7 are the only two indicators in a larger set based on the self-report method. Because they share a common method, scores on X_1 and X_7 may covary even if each indicator depends on a different factor. Specific variances of repeated measures variables may also overlap, a pattern called autocorrelation. Correlated errors are represented in CFA

model diagrams with the same symbol as for factor correlations, such as E_1 E_7 for the indicator pair X_1 and X_7.

Standard CFA models specify unidimensional measurement for all indicators. There are two different ways to specify multidimensional measurement, which also implies a non-standard CFA model: if any indicator is regressed on 2 or more factors (i.e., there are ≥2 pattern coefficients for the same indicator) or if correlated errors are specified. For example, adding the direct effect B → X_1 to the model of Figure 3 (b) would specify multidimensional measurement. There is controversy about allowing indicators to depend on more than one factor. On the one hand, some indicators may actually measure more than one domain. An engineering aptitude test with text and diagrams, for instance, may measure both verbal and visual-spatial reasoning. On the other hand, unidimensional models offer more precise tests of convergent validity and discriminant validity. A set of variables presumed to measure the same construct shows convergent validity if their inter-correlations are at least moderate in magnitude. In contrast, a set of variables presumed to measure different constructs shows discriminant validity if their inter-correlations are not too high. If $r_{XY} = .90$, for instance, then we can hardly say that variables X and Y measure distinct constructs, refer to (Kline, 2010) and (Kline, 2013).

The numerals (1) that appear in Figure 3 (b) next to paths that point to indicators or error terms are scaling constants that represent the assignment of a scale or metric. Because factors and error terms are not directly measured, they require a scale before the computer can generate statistical estimates for them. For example, the specification.

$$E_1 \rightarrow X_1 = 1.0 \tag{13}$$

In Figure 3 (b) scales the error term so that its variance is related to that of the unexplained variance in the corresponding indicator.

$$A \rightarrow X_1 = 1.0 \; and \; B \rightarrow X_4 = 1.0 \tag{14}$$

The specifications in the figure scale the factors such that their variances are related to that of the explained (common) variance of the corresponding indicator. Because the scales of

X_1 and X_4 are "borrowed" in order to scale the factors, each of these indicators is a reference variable. Assuming that indicators of the same factor have equally reliable scores, it is actually arbitrary which indicator is selected as the reference variable, but both factors and error terms must be assigned scales in CFA (and in SEM, too). Some SEM computer programs, such as LISREL and Mplus, scale factors and error terms automatically, but other programs may require the user to explicitly scale each of these variables. There are actually scaling constants in EFA, but computer procedures for EFA automatically assign these constants as part of the analysis.

Readers should not over interpret the labels "exploratory" and "confirmatory." It is true that EFA does not require a priori hypotheses about factor-indicator correspondence or even the number of factors. However, there are ways to conduct EFA in a more confirmatory mode, such as instructing the computer to extract a certain number of factors based on theory. Also, the technique of CFA is not strictly confirmatory. Specifically, it happens in perhaps most analyses that the initial restricted measurement model does not fit the data. In this case, the researcher typically modifies the hypotheses on which the initial model was based and specifies a revised model. The re-specified model is then tested again with the same data. The goal of this process is to "discover" a model with three proper-ties: It makes theoretical sense, it is reasonably parsimonious, and its correspondence to the data is acceptably close.

This is a good point to mention two critical issues in factor analysis. One is the factor indeterminacy problem, which is that hypothetical constructs can basically never be uniquely determined by their indicators. This means that although the results of a factor analysis might indicate that a particular measurement model is consistent with observed covariances, there may be nevertheless be other factor structures just as consistent with the same data. A more modern expression of the same idea refers to the problem of equivalent models, and for measurement models with multiple factors there are actually infinitely many equivalent versions. This is not a fatal fl aw of factor analysis but instead a characteristic of statistical modelling in general. As noted by Mulaik (1987), such techniques are best seen as hypothesis-generating methods that support inductive reasoning but do not produce definitive, incontrovertible results. The second critical issue concerns the

naming fallacy, or the false belief that the name assigned to a factor by a researcher means that the hypothetical construct is understood or even correctly labelled. Factor names are descriptions, not explanations, so we cannot assume that a particular factor label is necessarily the correct one. An example where the same two factors are assigned different labels by different researchers is presented later, but factor labels should be considered as hypotheses and not as substitutes for critical thinking.

3.5.7. Exploratory Factor Analysis

Summarized next is the typical sequence of additional decisions in EFA. These steps are sometimes iterative because results at a later phase may necessitate a return to an earlier step:

1. Select a method of factor extraction. The most basic choice is between principal axes factoring (PAF)—also known as common factor analysis—and principal components analysis (PCA). The PCA method is the default in some EFA computer procedures, such as SPSS, but this option is not always the best.
2. Decide how many factors to retain. There are two kinds of criteria for making this choice, theoretical and statistical. Of the two, theoretical criteria may result in less capitalization on sample-specific (chance) variation than statistical criteria.
3. Select a rotation method and interpret the factors. The goal of rotation is to enhance the interpretability of the retained factors. There are many different rotation methods, but the most basic choice is between some types of orthogonal rotation where the factors are independent or oblique rotation where the factors are allowed to covary. The default method in most EFA computer procedures is orthogonal rotation, but this option is not always ideal.

Factor Extraction Method

The difference between PAF and PCA—and the only difference—is the form of the data matrix analyzed (Thompson, 2004). The PCA method assumes that all indicator variance is common (shared) variance. The assumption is strict because it does not allow for specific variance or measurement error (see Figure 2); that is, the method assumes that the scores are perfectly reliable. Accordingly, all observed variance is analyzed in PCA. This means that

the correlation matrix analyzed by PCA has diagonal entries that all equal 1.0, which is literally all the observed variance in standardized form. The data matrix analyzed in PCA is thus called an unreduced correlation matrix. In contrast, the PAF method analyzes common variance only. This means that the diagonal entries of 1.0 in the correlation matrix are replaced in the PAF method by h2 statistics, or estimated communalities for each indicator. Suppose that the estimated communality for indicator X_3 is .75. In the correlation matrix, the 1.0 in the diagonal entry for X_3 will be replaced by .75 in PAF. All remaining diagonal entries of 1.0 are also replaced by the corresponding h^2 value for each of the other indicators, and in each case $h2 \leq 1.0.$ Thus, it is a reduced correlation matrix that is analyzed in PAF. When a covariance matrix is analyzed, the diagonal entries in PAF are replaced by the product of the sample variance and the communality estimate for each indicator. Because the PAF method analyzes common variance, it does not assume perfect score reliability.

Statistical procedures for PAF typically use an iterative method to estimate communalities where the computer derives initial estimates and then attempts to improve these estimates through subsequent cycles of calculations. The default initial estimate for each indicator is usually the squared multiple correlation (SMC) between that indicator and all rest. For example, if SMC = .60 for indicator X4, then 60% of the observed variance in X4 is explained by all the other indicators. However, sample correlations (and squared correlations, too) can be attenuated by measurement error, and iterative estimation takes account of this phenomenon. Sometimes in PAF it happens that iterative estimation fails, that is, the computer is unable to derive a final set of communality estimates. Iteration failure may be indicated by a warning or error message in the output. Any subsequent estimates in the rest of the output should be ignored. Another sign of trouble are Heywood cases, or estimates that are mathematically impossible, such as a structure coefficient >1.0 or a negative (<0) estimate of error variance. Solutions with Heywood cases are inadmissible and warrant no further interpretation. Some PAF computer procedures allow the user to increase the default limit on the number of iterations, which gives the computer more "tries" and may solve the problem. Some programs also accept user-specified initial communality estimates that replace the default initial estimate. Failure of iterative

estimation is more likely in small samples, when score reliabilities are low, or there are too few indicators of some factors. Communalities are not estimated in PCA, so the potential problems of iteration failure and Heywood cases are not encountered in this method.

The conceptual difference between PAF and PCA is that factors in PCA are estimated as composites, or weighted linear combinations of the indicators (i.e., total scores). How-ever, factors in PAF have the status of latent variables that are estimated taking account of measurement error. For example, the representation of the two-factor, six-indicator model in Figure 3 (b) is consistent with the PAF method because common variance is analyzed apart from unique variance, which corresponds to the error terms in the figure. Presented in Figure 4 is a depiction of a two-factor, six-indicator model analyzed in PCA. The factors in Figure 4 are each represented with hexagons, which some authors use in model diagrams to represent composites. The lines with single arrowheads that point from the indicators to the factors represent the fact that factors are estimated as weighted total scores across the indicators in PCA. From this perspective, indicators are the predictors in PCA, but the predictors in PAF are the factors (compare Figure 3 (b) and Figure 4). The dashed line in the symbol for a covariance in Figure 4 represents the possibility for an oblique factor rotation in PCA.

Because the PCA method analyzes observed variables only, some methodologists do not consider it to be a "true" method of factor analysis. Instead, PCA is described by those who hold this view as a mere data reduction technique that replaces a set of correlated variables with a smaller set of orthogonal composites, not as method for estimating latent variables. Others refer to the composites generated in PCA as "components" instead of "factors." There has been much debate in the literature about the relative merits of PCA versus PAF (e.g., Mulaik, 1992), some of it quite rancorous. It helps to know that PAF and PCA tend to generate similar solutions when scores on all indicators are very reliable, each indicator depends mainly on just one factor, all communalities are relatively high, and the sample size is large (e.g., Velicer & Jackson, 1990). Otherwise, the two methods can generate appreciably different estimates when applied to the same data. Results of PCA and PAF also tend to converge as more and more indicators are analyzed. This hap-pens because the number of diagonal elements as a proportion of all the elements in a data matrix decreases

as the number of variables increases (Thompson, 2004), and PCA and PAF analyze the same data matrix except for the diagonal entries. In SPSS, the initial factor solution is extracted using PCA even if the user requested PAF extraction for the final solution. In general, PAF is a better choice than PCA when not all score reliabilities are high (e.g., $r_{XX} > .80$).

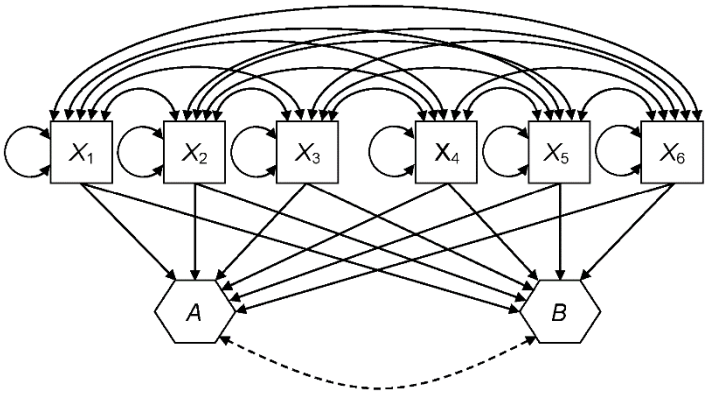

Figure 4 Conceptual representation of principal components analysis for 6 indicators and 2 composites (components).

Some other EFA extraction methods are briefly described next; see Mulaik (2009, chapter 7) for more information. The method of alpha factor analysis is a generalization of the internal consistency reliability coefficient (i.e., Cronbach's alpha) but in this case applied to indicators of different factors instead of to items within the same test. This method associates indicators with factors in a way that maximizes the internal consistency of construct measurement. The method of image analysis minimizes the chance that a factor will be defined mainly by a single indicator. This outcome is undesirable in factor analysis because single-indicator measurement is generally imprecise compared with multiple-indicator measurement. Image analysis works by minimizing the effect of specific variance on the results. In maximum likelihood factor analysis, the method of ML estimation is used to derive common factors that reproduce sample correlations among the indicators as close as possible. The method also generates statistical tests of parameter estimates, including the pattern coefficients and factor correlations. In contrast, these kinds of statistical tests are

not available in the PAF and PCA methods. In SPSS, the ML method is applied to correlation matrices only.

Number of Retained Factors

Sometimes the decision about the number of factors to retain is determined by theory, such as when a cognitive ability test battery is constructed in order to assess three different underlying domains. In this case, it makes sense to specify a three-factor solution. Even when theory indicates a specific number of factors, many researchers will nevertheless inspect a range of factor solutions, such as solutions with two, three, or four factors for the example just mentioned. Doing so not only evaluates the original hypothesis about the presence of three factors; it also allows for alternative explanations (i.e., two or four factors) to be tested. There are also various statistical criteria for determining the number of retained factors. It is best not to blindly apply these criteria because doing so tends to capitalize on chance variation in a particular sample. Instead, a researcher should inspect statistical criteria in light of extant theory.

This first statistical criterion is the default basis in most EFA computer procedures for determining the number of retained factors, but this method is not always the best choice. It is the eigenvalue >1.0 rule, also known as the Kaiser criterion or K1 rule after the educational statistician Henry F. Kaiser. Every extracted factor in EFA has its own eigenvalue, which is a measure of the proportion of variance in the indicators explained by that factor and is designated by the symbol λ. The first extracted factor tends to have the highest eigenvalue, and eigenvalues tend to successively decrease as additional factors are extracted. An eigenvalue is the sum of the squared structure coefficients across all the indicators. If λ = 1.0, then the amount of variance explained by that factor corresponds to the amount of information in one indicator, and if λ = 2.0, the explanatory power corresponds to the variability of two indicators, and so. The ratio λ/n_j where n_i is the number of indicators is the proportion of total variance explained by the associated factor. The sum of the eigenvalues for all possible factors equals the number of indicators, and the ratio of the sum of the eigenvalues across all retained factors divided by the number of indicators is the total pro-portion of variance explained by the factors as a set (Thompson, 2004). These proportions of explained variance are another type of variance-accounted-for effect size in EFA. The λ>1.0 rule thus requires that a factor explains at least one "unit" of information in terms of the indicators. However, note that the λ >1.0 rule applies to factors

extracted using the PCA method, not the PAF method. Indeed, it is a mistake to apply this rule when a reduced correlation or covariance matrix is analyzed (Fabrigar et al., 1999).

There are two problems with blindly following the λ >1.0 rule. First, sampling error affects the estimation of eigenvalues, which ignores that the fact that λ =1.01 versus λ =.99 leads to categorically different outcomes (i.e., retain vs. do not retain) under this rule. Second, factor solutions determined by the λ >1.0 rule tend to have too many factors but also occasionally too few factors (e.g., Velicer & Jackson, 1990). In general, the λ >1.0 rule is not a sound basis for deciding how many factors to retain. A variation is the Cattell scree test (after the psychologist Raymond B. Cattell), which is a visual heuristic that involves making a line graph where eigenvalues are plotted on the Y-axis for each of the total possible number of factors, which are represented on the X-axis. The graph is then visually inspected in order to locate the point where the drop in eigenvalues over successive factors levels out and from which the slope of the line is basically horizontal. The number of retained factors in this approach corresponds to the number of eigenvalues before the last substantial drop in the graph. A drawback of the scree test is that its interpretation can be rather subjective in that two different researchers can come to different conclusions after inspecting the same plot.

A more sophisticated method based on eigenvalues is that of parallel analysis, which involves the comparison of observed eigenvalues against those expected from random data. One way to conduct a parallel analysis is to use a computer procedure to randomize the scores for each variable in a raw data file. The randomized scores have the same distributional characteristics of the original scores, but their expected inter-correlations are about zero. Next, eigenvalues from the analysis of the randomized scores are compared against those from the original data. Factors are retained whenever the eigenvalues based on the original scores are greater than the eigenvalues for the corresponding factors based on the randomized scores. Suppose that the eigenvalues for the first three factors based on analyzing the original scores are, respectively, 5.525, 2.350, and 1.026. The corresponding eigenvalues based on the randomized scores are 2.770, 1.850, and 1.332. This pattern suggests that first two factors should be retained but not the third because 1.332 (based on the randomized scores) is greater than 1.026 (based on the original scores). Results of

computer simulation studies about parallel analysis have generally been favourable (e.g., Crawford et al., 2010), but Fabrigar et al. (1999) noted that the decision to retain factors can be rather arbitrary when the pair of eigenvalues for a particular factor are very similar.

Most researchers retain fewer factors than the number of indicators, and such solutions typically explain some proportion of the total variance of the indicators, but usually not all of it. This also means that the factors will not perfectly explain the sample correlations (i.e., the data). Many EFA computer procedures optionally print predicted correlations also called reproduced correlations—and residual correlations. The former are the values of predicted Pearson correlations for the set of indicators, given the factor solution, and the latter are the differences between the observed and predicted correlations for each indicator. The lower the residuals in absolute value, the better the explanatory power of a factor solution. There is no absolute cut-off for interpreting residual correlations, but a better result is indicated if all absolute residuals are < .05. However, absolute residuals >.10 suggest poor prediction of the corresponding correlation.

It is generally better in EFA to extract too many factors, or over factoring, than to retain too few, or under factoring. This is because there tends to be substantial error in the results with under-factoring. For example, estimation of pattern coefficients of indicators that actually depend on a factor may complicate accurate estimation of coefficients of other indicators that measure retained factors. Two factors that are really distinct may be merged in a model with too few factors, which may complicate interpretation of the underlying latent variables (Fabrigar et al., 1999). Over-factoring is not without penalty, including the interpretation of factors that do really correspond to hypothetical constructs in the population. Over factoring is also a potential drawback of the $\lambda > 1.0$ rule.

Method of Rotation

The option for rotation does not apply when only a single factor is retained; otherwise, rotation is part of just about all analyses of multifactor models in EFA. The initial factor solution is often difficult to interpret. This is because the structure coefficients for the associations between the first factor and all indicators tend to uniformly high, and patterns of correlations with the remaining factors may not be very distinct. The goal of rotation is to

make the meaning of the factors more obvious to the researcher. It works by reweighting the initial factors (i.e., the factor axes are shifted) according to statistical criteria that vary with the particular method of rotation. The desired outcome is a rotated solution that exhibits simple structure where each factor explains as much variance as possible in non-overlapping sets of indicators. That is, the structure coefficients for the rotated factors should head toward either 0 or 1.0 in order to make the associations between factors and indicators more distinct. Theoretically, there are an infinite number of possible factor rotations for a given solution. In practice, either the researcher specifies a rotation option in a computer procedure for EFA or the computer will employ its default method.

There are two main classes of rotation methods in EFA. In orthogonal rotation, the rotated factors are all uncorrelated just as they are in the initial solution. Consequently, values of the standardized pattern coefficient and the structure coefficient for each indicator are equal. The most widely used rotation method of any kind is Kaiser's varimax rotation, which is also the default in SPSS. Varimax rotation maximizes the variance of the structure coefficients (i.e., it pushes them toward 0 or 1.0) for each factor, which tends to *(a)* limit the number of indicators with high correlations with that factor and *(b)* evenly distribute the indicators across the factors. This method is used so often because it generally yields simple structure in perhaps most EFA studies where the factors are uncorrelated (Thompson, 2004). A related method is quartimax rotation, which maximizes the variance of the structure coefficients for each indicator. This tends to yield a factor solution with a general factor that correlates highly with most indicators and lesser factors each associated with different subsets of indicators. The method of equamax rotation offers a kind of "comprise" in that it maximizes a weighted function of the criteria from the varimax method and the quartimax method. Selection among these alternatives for orthogonal rotation should be guided by the researcher's hypotheses about the anticipated model.

The assumption of all orthogonal rotation methods is that the underlying constructs are uncorrelated, but this hypothesis is not always defensible (Fabrigar et al., 1999). For example, it makes little sense that various latent cognitive abilities, such as verbal reasoning, visual-spatial reasoning, and memory, would be unrelated to each other. It seems just as implausible to presume that certain latent affective domains, such as anxiety and

depression, are independent. Methods of oblique rotation allow the factors to covary; that is, they estimate measurement models with correlated factors. Note that the specification of oblique rotation does not somehow "force" the factors to co-vary. Instead, these methods estimate factor correlations, given the data, so these estimates are "allowed" to be close to zero, if such estimates are consistent with the data. However, if an estimated factor correlation is extremely high (e.g., >0.9), then the two factors are clearly not distinct (i.e., there are too many factors). Also, the pattern coefficients and the structure coefficient for the same indicator are typically unequal when the factors covary. The assumption of all orthogonal rotation methods is that the underlying constructs are uncorrelated, but this hypothesis is not always defensible (Fabrigar et al., 1999). For ex-ample, it makes little sense that various latent cognitive abilities, such as verbal reasoning, visual-spatial reasoning, and memory, would be unrelated to each other. It seems just as implausible to presume that certain latent affective domains, such as anxiety and depression, are independent. Methods of oblique rotation allow the factors to covary; that is, they estimate measurement models with correlated factors. Note that the specification of oblique rotation does not somehow "force" the factors to covary. Instead, these methods estimate factor correlations, given the data, so these estimates are "allowed" to be close to zero, if such estimates are consistent with the data. However, if an estimated factor correlation is extremely high (e.g., >0.90), then the two factors are clearly not distinct (i.e., there are too many factors). Also, the pattern coefficients and the structure coefficient for the same indicator are typically unequal when the factors covary.

Promax rotation is probably the most widely used oblique rotation method. It usually begins with a varimax-rotated solution and then raises the pattern coefficients to a higher power κ (kappa), which tends to force near-zero coefficients to approach zero faster than coefficients further from zero (Mulaik, 2009). The procedure then generates least squares estimates of the target matrix just described by allowing the factors to covary. A parameter of promax rotation is κ, or the power to which coefficients in the target matrix are raised. Values of κ usually range from 1 through 4, where higher values permit higher absolute estimates of factor correlations that are consistent with data. The default in SPSS is $\kappa = 4$, and it is usually unnecessary to change this value. Another oblique method is direct oblimin

rotation, which is a member of a family of oblimin methods that generally minimize the variance of the pattern coefficients across different factors while estimating factor correlations. The parameter for direct oblimin rotation in SPSS is referred to as δ (delta), which ranges from negative to positive in value up to a maximum of .80 (i.e., δ ≤.80). Lower values of δ tend to decrease absolute estimates of factor correlations and higher values result in just the opposite. The default in SPSS is δ=0, and is rarely necessary to specify a different value.

A potential drawback of oblique rotation is that interpretation of the factors may be more difficult. One reason is that there are two sets of standardized coefficients, pattern and structure, and it can happen that and their values for the same indicator are quite different. For example, if an indicator's structure coefficient is about zero but its pattern coefficient is not, then a suppression effect is indicated. This means that the indicator contributes to the factor indirectly by changing the relations of other indictors to the same factor (Thompson, 2004). Another sign of suppression is when the signs of the pattern coefficient and structure coefficients for the same indicator are different. A second complication of oblique rotation is that there are few guidelines for applied researchers concerning the specification of values of parameters for the promax and oblimin methods other than the default values that would be suitable for a particular set of indicators (Costello & Osborne, 2005).

There are many other rotation methods in EFA, and it can be difficult to decide which method is best. There is also some degree of trial and error in their use. For example, it could happen in a particular analysis that a promax oblique rotation generates results that are easier to interpret than an oblimin oblique rotation. However, if the results are dramatically different after application of different rotation methods, then there may be little basis to pick one solution or the other when no replication sample is available. How-ever, a robust underlying measurement model with simple structure assessed with psycho-metrically sound indicators should be detected by different rotation methods. See Mulaik (2009, Chapters 10–12) for more information about factor rotation in EFA.

EFA Empirical Example

The first edition of the Kaufman Assessment Battery for Children (KABC-I; Kaufman & Kaufman, 1983) is an individually-administered cognitive ability test for children ages 2½–12½ years old. The test's authors claimed that the KABC -I's eight subtests measure two factors. The three tasks believed to reflect sequential processing all require the correct recall of auditory stimuli (Number Recall, Word Order) or visual stimuli (Hand Movements) in a particular order. The other five tasks Gestalt Closure, Triangles, Spatial Memory, Matrix Analogies, and Photo Series are supposed to measure more holistic, less order-dependent reasoning, or simultaneous processing. Each of these tasks requires that the child grasps a gestalt but with somewhat different formats and stimuli that all involve visual stimuli, refer for more details from (Kline, 2010) and (Kline, 2013).

The following conclusion was drawn bases on EFA after considering the above discussed methodology. Overall, the EFA results are consistent with a two-factor model, but the Hand Movements task of the Sequential Processing scale is problematic because it seems to belong more with the tasks on the Simultaneous Processing scale. Results of CFA for these data described in the next section yield even more precise results concerning this particular indicator.

3.5.8. Confirmatory Factor Analysis

As in EFA, the decision sequence for CFA begins with construct definition, indicator selection, and a sampling plan. Before collecting the data, the researcher should also specify the restricted measurement model to be analyzed in CFA and then check whether it is identified. The former requires specification of the number of factors in the model and the pattern of indicator-factor correspondence. The hypothesis of unidimensional measurement requires that there is a single pattern coefficient for each indicator and that there are no correlated errors (e.g., Figure 3 (b)), that is, the model is a standard CFA model. CFA models described in the literature are standard models. Specification of multidimensional measurement is an option, but doing so requires a substantive rationale in order to allow indicators to depend on more than one factor or the specification of error correlation between a pair of indicators. Such nonstandard CFA models are more complex than standard models, and they are not always identified, an issue considered next, (Kline, 1994, 2010 & 2013).

Identification Requirements

A CFA measurement model is identified if it is theoretically possible for the computer to derive a unique estimate of every model parameter. The word "theoretically" emphasizes identification as a property of the model and not of the data. For example, if a model is not identified, then it remains so regardless of the sample size (N = 100, 1,000, etc.).

Therefore, models that are not identified must be re-specified; otherwise, attempts to analyse them may be fruitless. There are two necessary but insufficient requirements for identification: (1) Every factor and error term must be assigned a scale, and (2) the model degrees of freedom must be at least zero ($df_M > 0$) The first requirement just mentioned was discussed earlier and is represented in diagrams of CFA models with the scaling constant"1" (e.g., Figure 3 (b)).

The quantity df_M is the difference between the number of observations available in the analysis and the number of model parameters, which in CFA are the pattern coefficients, factor variances and covariances, and error variances and covariances. The number of observations is not the sample size. Instead, it is literally the number of entries in the data matrix in lower diagonal form where only the unique values of correlations or covariances are reported in the lower-left-hand side of the matrix. The number of observations is calculated as $n_i(n_i + 1)/2$, where ni is the number of indicators, not the sample size. For example, if there are n_i = 4 indicators in a CFA model, then the number of observations is 4(5)/2, or 10. This count (10) equals the total number of diagonal and unique off-diagonal entries in the data matrix for 4 variables. With n_i = 4, the greatest number of parameters that could be estimated by the computer is 10. Fewer parameters can be estimated in a more parsimonious model, but not > 10. Also, the number of observations has nothing to do with sample size. If four indicators are measured for 100 or 1,000 cases, the number of observations is still 10. Adding cases does not increase the number of observations; only adding indicators can do so.

In practice, researchers should analyze models with positive degrees of freedom ($df_M \geq 0$). This is because identified models with no degrees of freedom will perfectly fit the data, that is, all residual correlations will equal zero. When $df_M = 0$, the model is just as complex as

the data to be explained. Such models are uninteresting because they test no specific hypotheses. Models where $df_M > 0$ generally do not have perfect fit. This is because dfM> 0 allows for the possibility of model-data discrepancies. Thus, retained models with greater degrees of freedom have withstood a greater potential for rejection. The idea underlies the parsimony principle: given two models with similar fit to the same data, the simpler model is preferred, assuming that it is theoretically plausible. Thus, the goal is thus to find a parsimonious measurement model with acceptably close fit to the data.

Additional identification requirements for standard CFA models concern the minimum number of indicators for each factor. A single-factor standard model requires at least three indicators in order to be identified. However, one-factor models with just three indicators have no degrees of freedom, so their fit to the data will be perfect, so in practice, a one-factor model should have ≥4 indicators. A standard model with ≥2 factors requires at least two indicators per factor in order to be identified. However, the analysis of CFA models where some factors have just two indicators is potentially problematic, so at least three indicators per factor is recommended.

The case concerning identification for nonstandard CFA models is more complicated. This is because unlike standard models, nonstandard CFA models that satisfy all the requirements just described are not always identified. In particular, specifying that an indicator depends on more than a single factor or that a pair of error terms is correlated is possible only if certain additional requirements are met. These extra requirements are summarized in the form of identification heuristics for determining whether a nonstandard model is identified (e.g., Kenny, Kashy, & Bolger, 1998; Kline, 2010, Chapter 6), but these heuristics are not always straightforward to apply for complex models with multiple correlated errors or indicators with ≥2 pattern coefficients. For example, in order for a model with error correlations to be identified, each factor must have at minimum number of indicators whose errors are uncorrelated, but this minimum number is either two or three depending on patterns of error correlations and pattern coefficients among the other indicators. There are similar requirements for each pair of factors and for each indicator in a nonstandard model. The specification of a single error correlation or that an indicator measures two factors in a CFA model that is otherwise standard may not cause a problem. This is another reason to

specify an initial CFA model that is parsimonious: Simpler models are less likely to run into problems concerning identification.

Parameter Estimation

The default method in CFA is ML, which in SEM analyzes covariance matrices only and simultaneously estimates all model parameters in an iterative algorithm. Computer procedures for ML estimation often begin iterative estimation by generating default initial estimates of certain model parameters known as start values. For example, the EQS program for SEM (including CFA) (Bentler, 2006) assumes in the first iteration that all unstandardized pattern coefficients in CFA models equal 1.0. However, if default start values are grossly inaccurate, then iterative estimation may fail to converge. Fortunately, most SEM computer tools allow the user to specify start values other than the program's default values. Better initial estimates may lead to a converged solution; see Kline (2010, p. 263) for guidelines on how to specify start values in CFA. If estimation converges successfully, it is still necessary to carefully look through the estimates for Heywood cases, just as in EFA when using PAF extraction.

The method of ML estimation assumes multivariate normality, and the method is not robust against violations of this assumption. This means that it is necessary to carefully screen the raw data and deal with problems, such as extreme outlier scores or severely non-normal univariate distributions that contribute to multivariate non-normality. Kline (2010, chapter 3) describes how to screen the data and prepare a "proper" matrix summary for ML estimation. Many SEM computer tools can optionally use a corrected nor-mal theory method, which uses ML estimation to generate parameter estimates that are then corrected for the degree of skew or kurtosis in the data. These corrected methods require input of a raw data fi le, not a matrix summary. There are other, more specialized estimation methods for severely non-normal data or for indicators that are not continuous variables, such as when items are specified as indicators in CFA models instead of scales. Some options for analyzing items-as-indicators in CFA are described in the next section.

Evaluation of Model Fit

There are two main classes of statistics in SEM that evaluate the correspondence between model and data, model test statistics and approximate fit indexes. The former are statistical tests of whether the covariance matrix implied by the researcher's model is close enough to sample covariance matrix that the differences might reasonably be considered as due to sampling error. Most model test statistics are scaled such that higher values indicate increasingly poor model-data correspondence. Thus, it is a statistically significant result (e.g., p < .05) that indicates problematic model–data correspondence. This logic is "backward" compared with most statistical tests where rejection of the null hypothesis supports the research hypothesis. But in SEM (and CFA, too) it is the lack of statistical significance (e.g., p> .05) that supports the researcher's model. The most widely reported test statistic is the model chi-square, χ^2_M , with degrees of freedom that equal df_M, the model degrees of freedom. The statistic χ^2_M assumes multivariate normality, which is also required in ML estimation.

In small samples, it can happen that the power of the model chi-square test is so low that it is difficult to correctly reject a false model (e.g., MacCallum et al., 1996). In very large samples, it can happen that χ^2_M is statistically signifi cant even though the magnitudes of model-data discrepancies are slight. For this reason, researchers in the past tended to ignore the results of the model chi-square test even in samples that were not very large. However, this practice is now viewed by more and more methodologists as excessively lax (e.g., Barrett, 2007). A better alternative is to consider a statistically significant result as providing preliminary evidence against the model that must be further diagnosed. Likewise, a model chi-square test result that is not statistically significant does not automatically lead to the decision to retain the model, especially if the sample size is not very large. Further evaluation of a model's fit to the data is also needed in this case (Kline, 2010).

The outcome of an approximate fit index is not the dichotomous decision to reject or retain a null hypothesis. Instead, these indexes are intended as continuous measures of model-data correspondence. Some approximate fit indexes are scaled such that lower values indicate closer model-data correspondence, but most are scaled so that it is higher values that suggest better fit. And the metrics of some approximate fit indexes are more-or-less standardized so that their range is 0–1.0 where a value of 1.0 indicates the best fit.

Philosophically, approximate fit indexes are consistent with the view that truly correct models may not exist. That is, basically all statistical models are probably wrong to some degree because they are imperfect reflections of a complex reality. Instead, models are approximation tools that help researchers to structure their thinking about a phenomenon of interest. In contrast, a model test statistic is more analogous to a smoke detector: When the alarm sounds, there may or may not be a fi re (serious model-data discrepancy), but it is prudent to treat the alarm seriously (conduct more detailed diagnostic evaluation of fit).

There are dozens of different approximate fit indexes, but most break down into a few basic categories. Briefly, absolute fit indexes are generally interpreted as proportions of the covariances in the sample data matrix explained by the model. However, explanatory power at the level of data matrix has little to do with whether the model accounts for relatively high proportions of the variance in the indicators. Incremental fit indexes indicate the relative improvement in fit of the researcher's model compared with a statistical baseline model where it is assumed that all observed variables are uncorrelated. But the assumption of zero covariances among indicators of a measurement model is implausible in most cases. Parsimony-adjusted fit indexes includes in their equations a built-in "penalty" for complexity related to the value of df_M. (Recall that more parsimonious models have higher degrees of freedom.) And predictive fit indexes estimate model fit in hypothetical replication samples randomly drawn from the same population, but most applications of CFA do not call for this type of fit index.

Based on the results of some computer simulation studies conducted in the 1980s and 1990s about the behavior of approximate fit indexes under varying data and model conditions, many researchers relied on a series of rules of thumb or threshold values of approximate fit indexes that supposedly indicated "good" fit of the model to the data. An example of a threshold for the hypothetical XYZ index scaled from 0–1.0 would be, if XYZ > .90, then conclude "good" fit. At the same time that researchers increasingly relied on threshold values for approximate fit indexes, they also tended to ignore model test statistics. Unfortunately, results of more recent simulation studies indicate that (1) the accuracy of thresholds depend on the particular type of structural equation model studied, (2) expected values of approximate fit indexes have little relation to their threshold values when

distributional assumptions are violated, and (3) there is little direct relation between values of fit statistics of any type (including model test statistics) and the degree or type of misspecification (e.g., Millsap, 2007). The point just mentioned explains why researchers should also provide more specific information about model fit. Perhaps the best way to do is to report the matrix of residual correlations, which say something about model fit at a more fundamental level than summary fit statistics (Kline, 2010). As in EFA, absolute residual correlations > .10 in CFA suggest poor explanation of the observed correlation between that pair of indicators. Briefly described next are approximate fit indexes the values of which should be reported in most analyses; see Kline (2010, chap. 8) for more information. Threshold values are given for each index, but readers should not reify these values in view of the issues just raised.

The Steiger-Lind root mean square error of approximation (RMSEA) is a parsimony corrected index that in computer output is usually reported with a 90% confidence interval, which takes account of sample size. Unlike χ^2_M, the RMSEA theoretically follows a noncentral chi-square distribution that allows for a certain degree of discrepancy between population and sample models. The best result is RMSEA = 0, and higher values indicate increasingly worse fit of the model to the data. If the value of the upper bound of the confidence interval based on the RMSEA exceeds .10, then problematic model-data correspondence may be indicated. The Bentler comparative fit index (CFI) is an incremental fit index that measures the relative improvement in fit of the researcher's model over that of a baseline model that assumes uncorrelated indicators. Like the RMSEA, the CFI allows for some discrepancy between population and sample models. Values of the CFI range from 0–1.0 where 1.0 is the best result. The standardized root mean square residual (SRMR) is a measure of the mean absolute residual correlation, so values close to 0 are a better result. Ideally, the value of the CFI should exceed .95 or so, and the value of the SRMR should be < .10.

Model Respecification

If the fit of an initial CFA model to the data is poor, then the analysis enters the respecification phase where alternative models are generated and then fitted to the same data matrix. There are two general types of respecification options. The first concerns the

correspondence between indicators and factors. The basic possibilities here include the respecifications that an indicator (a) loads on an additional factor, (b) depends on a different factor, or (c) shares has an error correlation with another indicator, all compared with the original model. The second category for respecification concerns the factors. For example, the researcher may have specified the wrong number of factors. Poor discriminant validity as evidenced by very high factor correlations may indicate that the model has too many factors. On the other hand, poor convergent validity within sets of indicators of the same factor suggests that the model may have too few factors.

Respecification should be guided by the same theoretical and empirical bases that supported the specification of the original model. At the same time researchers often inspect certain types of output that may inform respecification. Inspecting the pattern of residual correlations may shed light on the question of indicator-factor correspondence. Suppose that indicator X_4 is specified to measure factor A but (a) the residual correlations between it and all the indicators of factor B are large and positive and (b) the pattern coefficient for the $A \rightarrow X_4$ is reasonably large. This pattern suggests that X_4 may measure both factors. A different possibility when there is a large residual correlation for a pair of indicators is that their specific variances overlap, or their error terms covary.

Most SEM computer tools can optionally print values of modification indexes, which are interpreted as chi-square statistics with a single degree of freedom, or $\chi^2(1)$. A modifi cation index estimates the amount by which the overall model chi-square would decrease if a previously fixed to zero parameter were freely estimated, which adds the corresponding effect to the model. Suppose for a two-factor model that X_4 is specified to measure factor A but not factor B. In this model, the parameter $B \rightarrow X_4$ is specified to equal 0. If the modification index for this parameter equals 5.50, then (a) it is estimated that the value of χ^2_M will decrease by 5.5 points if the parameter $B \rightarrow X_4$ were added to the model and (b) the amount of this decrease would be statistically significant because $\chi^2(1). = 5.50$, $p > .05$. However, the researcher must avoid respecifying the model based solely on modification indexes. This is because these statistics capitalize heavily on sample specific variation, and respecification that blindly chases modification indexes is unlikely to lead to a true model (e.g., Silvia & MacCallum, 1988).

A related statistical test that is more generally useful for theoretically-based respecification is the chi-square difference test based on the $\chi_D^2 \quad (df_D)$ statistic, which is the difference between the χ_M^2 statistics for two hierarchically-related models fitted to the same data χ_M^2 where df_D equals the difference between the df_M values for the two models. Two CFA models are hierarchical if one is a subset of the other, that is, a simpler version formed by dropping ≥ 1 parameters from a more complex model. Assuming that the more complex model fits the data, the result of the difference test indicates whether the relative fit of the simpler models is statistically worse than that of the more complex model. If not, then dropping the parameters to form the simpler model may not appreciably worsen overall fit; otherwise, the more complex model is favored. In CFA, the chi-square difference test is often used to compare, say, a two-factor solution versus a simpler one-factor model. If the fit of the more parsimonious one-factor model is not appreciably worse than that of the more complex multifactor model, then the one-factor model is preferred.

Revised CFA Empirical Example

This example was application of CFA on the EFA empirical example discussed earlier. The details of this application can be found in Kelin (2010) and Kelin (2013). The following conclusions were drawn with regards to CFA in comparison to EFA. For this example, the fit of the EFA model was better than the fit of the CFA model to the same data. This is not unexpected because the EFA model allows each indicator to load on both factors, but the CFA model specifies unidimensional measurement. This is one reason why the specification of a CFA model based on EFA outcomes and analyzed with the same data may lead to the rejection of the CFA model (van Prooijen & van der Kloot, 2001). That is, CFA does not generally "confirm" the results of EFA. But the CFA results made more obvious the sources of poor fit, which concern mainly the three indicators of the sequential processing factor.

CFA Model Equations with Item Intercepts

Measurement model per item (numbered) for the subject s (Kline, 2013):

Steps:

- You decide how many factors and whether each item loads (loading then estimated) or not.
- Standardized loadings are the slopes in the a correlation metric (and $Std\ Loading^2 - reliability$).
- Standardized loadings are the correlation metric (and $Loading^2 - reliability$).
- Intercepts (μ) are expected value of Y (item) when all factors (X's) are 0 (misfit).

Derivations steps as explained above are as follows;

$$x_{1s} = \mu_1 + \lambda_{11}F_{1s} + 0F_{2s} + e_{1s} \quad (15)$$

Similary

$$x_{2s} = \mu_2 + \lambda_{21}F_{2s} + 0F_{2s} + e_{2s}$$

$$x_{3s} = \mu_3 + \lambda_{31}F_{3s} + 0F_{2s} + e_{3s}$$

$$x_{4s} = \mu_4 + \lambda_{41}F_{4s} + 0F_{2s} + e_{4s}$$

$$x_{5s} = \mu_5 + 0F_{1s} + \lambda_{52}F_{5s} + e_{5s}$$

$$x_{6s} = \mu_6 + 0F_{1s} + \lambda_{62}F_{6s} + e_{6s}$$

$$x_{7s} = \mu_7 + 0F_{1s} + \lambda_{72}F_{7s} + e_{7s}$$

$$x_{8s} = \mu_8 + 0F_{1s} + \lambda_{82}F_{8s} + e_{8s}$$

The equation predicting each item resembles a linear regression model;

$$Y_{is} = B_{0i} + B_{1s}X_{1s} + B_{2s}X_{2s} + e_{is} \quad (16)$$

Under the above setup, you decide the number of factors and whether each item loads or not (confirmation). Unstandardized loadings (λ) are the slopes of regression of the response (Y) on the factor (X). Standardized loadings are the slopes in a correlation metric (and standard loadings 2 = reliability). Intercepts (μ) are expected values of Y (item) when all factors $(X\ 'S)$ are 0 (no misfits).

Expressing the CFA Model in Matrices: Factor Loadings

If we put our loadings into a matrix **Λ** (size *p* items by *m* factors) (Kline, 2010 & 1994)

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \end{bmatrix} \tag{17}$$

Expressing the CFA Model in Matrices: Unique Variances

$$\Psi = \begin{bmatrix} \psi_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_2^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \psi_3^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \psi_4^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_5^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \psi_6^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_7^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \psi_8^2 \end{bmatrix} \tag{18}$$

Expressing the CFA Model in Matrices: Factor Covariances. If we put our factor covariances into a matrix **Φ** (size *m* factors by *m* factors):

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{bmatrix} \tag{19}$$

the result of the CFA model then predicts the observed covariance matrix of the items by:

$$\Sigma = \Lambda\Phi\Lambda' + \Psi \tag{20}$$

Where:

$$\Sigma = \begin{bmatrix} \lambda_{11}^2 + \psi_1 & \cdots & \lambda_{11}\phi_{12}\lambda_{62} \\ \vdots & \ddots & \vdots \\ \lambda_{11}\phi_{12}\lambda_{62} & \cdots & \lambda_{62}^2 + \psi_6 \end{bmatrix}$$

Examples of CFA Model Predictions

Consider having the following equations, (Kline, 2010)

$$F_1 \ by \ X_1 - X_4 \ , \tag{21}$$

$$F_2 \ by \ X_5 - X_8$$

Two items from same factor (room for misfit). We obtain two types of solutions as result. Which are

Unstandardized solutions:

$$\text{Covariance } (x_1, x_4) \ = \ \lambda 11 * Var(F_1) * \lambda_{41} \tag{22}$$

Standardized solutions:

$$\text{Correlation } (x_1, x_4) = \ \lambda_{11} * \lambda_{41} \ (std \ loadings). \tag{23}$$

Only reason for $cor(x_1, x_4)$ is common factor (local independence). Variances are additive (and will be reproduced correctly):

$$Var(X_1) \ = \ (\lambda 11^2) * Var(F_1) \ + \ Var(e_1) \tag{24}$$

note imbalance of λ_2 and e.

CFA Model Identification: Creating a Scale for the Latent Variable

The illustration below is a CFA model with factor means and item intercepts. But some of these values will have to be restricted for the model to be identified.

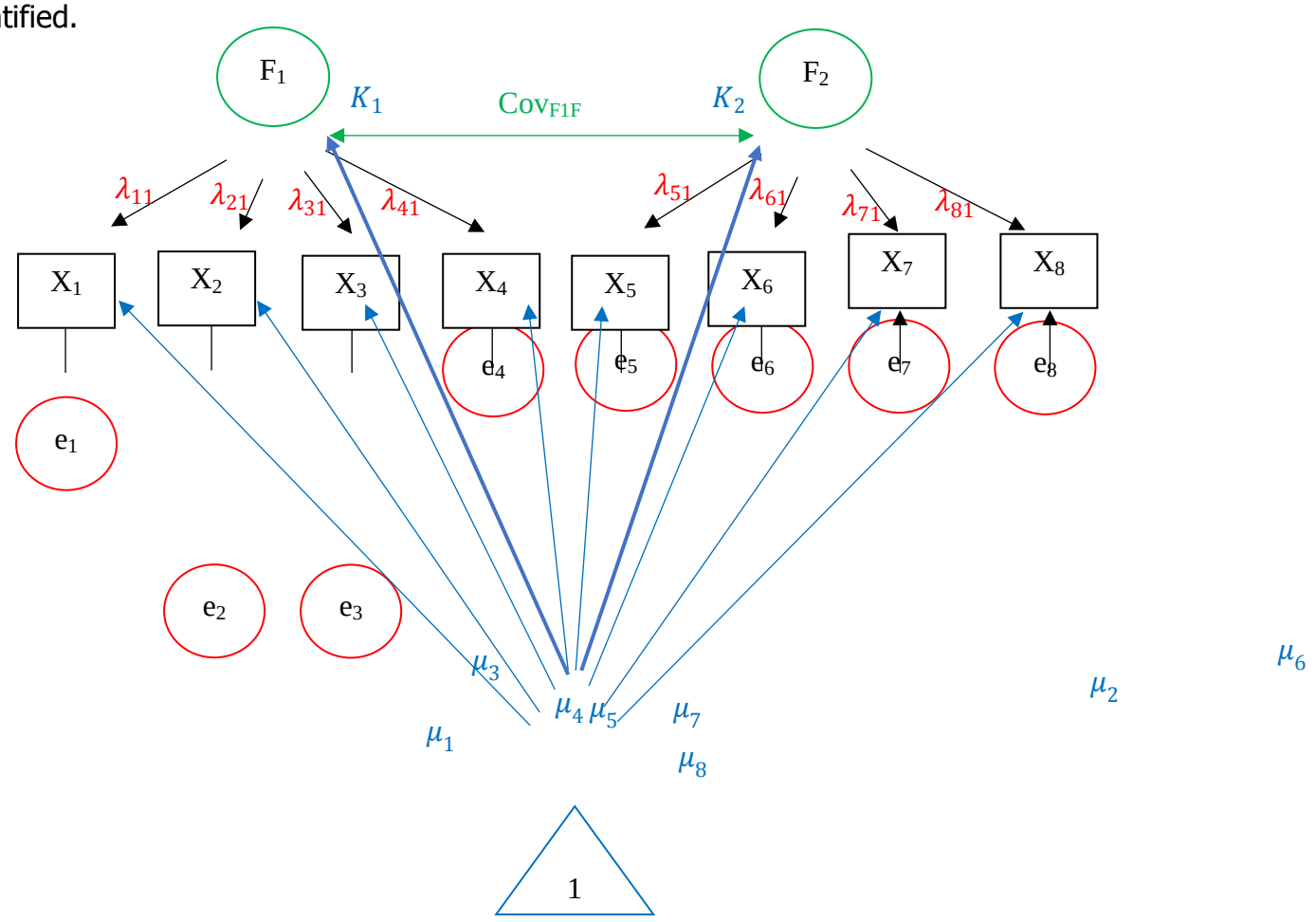

Figure 5 Demonstration of CFA model, showing the measurement model and the structural model.

There exist two models from the CFA model with their respective parameters from the above demonstration, which are;

Measurement Model

- $\lambda's = Factor\ loading$
- $e's = Error\ variances$
- $\mu's = Item\ intercepts$

Structural Model

- $F\text{'}s = Factor\ variances$
- $Cov = Factor\ covariances$
- $K\text{'}s = Factor\ means$

Note, some of the values indicated will have to be restricted for the model to be identified. The Big Conceptual Difference between PCA and EFA is that In PCA, we get components that are outcomes built from linear combinations of the items but in EFA, we get factors that are thought to be the cause of the observed indicators.

Examples of CFA Model Identification

CFA model identification requires a latent variable, which needs a scale (mean and variance. There are two options of how to create the scale required, see illustration Figure 12. They are as follows;

- Option 1: Create a scale using a marker item :
    - Fixing one loading to 1; factor is scaled as reliable part of that marker item.
    - E.g. Loading =0.9, variance=16. $Var(F_1) = 0.9\text{^}2 * 16 = 12.93$
- Option 2: Fix factor variance to 1
    1. Factor is interpreted as z-score.
    2. This can't be used in other models with higher-order factors

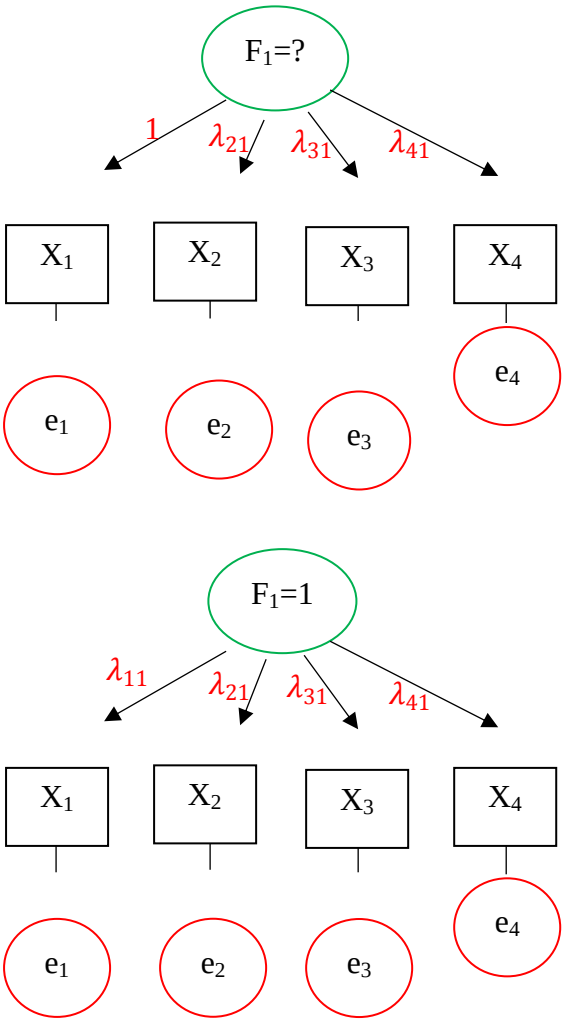

Figure 6 Types of CFA Model Identification.

3.5.9. Other Issues on Factor Analysis

Considered next are some additional issues and analysis options in factor analysis.

Items as Indicators

Likert-scale items are not generally continuous variables, and Pearson correlations may not be the best measure of association for them. This is especially true for items with binary response formats (e.g., true-false). Results of some computer simulation studies indicate that estimates from standard estimation methods for continuous variables may be

inaccurate when the indicators are binary or ordinal variables. These simulation studies generally assume a true population measurement model with continuous indicators. Within generated samples, the indicators are categorized to approximate data from non-continuous variables. Bernstein and Teng (1989) found that when there is only a single factor in the population but the indicators have few categories, one-factor measurement models tend to be rejected too often. That is, categorization can spuriously suggest the presence of multiple factors. DiStefano (2002) found that ML parameter estimates and their standard errors were both generally too low when the data analyzed are from categorical indicators, and the degree of negative bias was higher as distributions became increasingly non-normal.

There are special estimation methods for analyzing ordinal indicators, such as items, available in some SEM computer tools, including EQS, LISREL, and Mplus (see Chapter 7 on Item Response Theory). These methods do not solely rely on ML estimation. Instead, they feature alternative estimators based on weighted least squares methods (e.g., Kline, 2010, chap. 7). These methods may be applied to a matrix of tetrachoric correlations or polychoric correlations, not Pearson correlations. A tetrachoric correlation is for two dichotomous variables, and it estimates what the Pearson r would be if both variables were continuous and normally distributed. A polychoric correlation is the generalization of the tetrachoric correlation that estimates r for ordinal variables with ≥ 2 levels. Some computer procedures, such as the PRELIS program of LISREL, can export polychoric or tetrachoric correlation matrices for analysis in a different program. Another option is to analyze parcels instead of items. A parcel is a total score across a set of homogenous items each with a Likert-type format. Parcels are generally treated as continuous variables, and analyzing a Pearson correlation matrix based on parcels is not problematic. The score reliability of parcels (total scores) tends to be greater than that for the individual items. However, it is critical that the items in each parcel are homogeneous, or unidimensional; otherwise, the results may be misleading. There are also different ways to parcel items, including random assignment of items to parcels and groupings of items based on rational grounds, and the choice can affect the results; see T. Little, Cunningham, Shahar, and Wi-daman (2002) for more information.

Another complication of item-level factor analysis is that easy or frequently endorsed items tend to form factors that are distinct from other factors made up of harder or less commonly endorsed items (Nunnally & Bernstein, 1994). Such factors may reflect differential response base rates more than they mirror substantive latent variables. In this case, a multifactor solution would be spurious, rendering subsequent interpretations incorrect. The researcher should inspect the response of items that form different factors. If these rates vary systematically over the factors, then the results may be due more too statistical than substantive bases (O'Connor, 2000 & 2011).

Factor Scores

Kelin (2013) discusses that when raw data are analyzed, it is possible to calculate factor scores for each case. Because factors are not directly measured but instead through their indicators, such scores are only estimates of cases' relative standings on the factors. There is more than one way to calculate factor scores, however, and although scores derived using different methods tend to be highly correlated, they generally do not all yield identical rank orderings of the cases. For example, given structure coefficients, the technique of multiple regression can be used to derive estimated factor scores that are weighted combinations of the indicators and the factor. The weights derived in regression are those that lead to the closest correspondence between the factors and the estimated factor scores. Other methods for oblique solutions constrain the correlations among the factor scores to match those of the corresponding factors. Given that there is more than one way to derive estimated factor scores, Bollen's (1989) perspective on this matter is relevant: Researchers should probably refrain from making too fi ne a comparison on estimated factor scores. See Grice (2001) for more information about factor scores.

Factor scores are calculated in EFA more often than in CFA. One reason is that scores derived in EFA tend to be used as either predictors or outcomes in subsequent analyses that involve other variables. In contrast, factors in structural equation models can be represented as either predictors or outcomes of other variables in the model, latent or observed. Such models are not CFA models—they are actually structural regression models—but it is no special problem in SEM to estimate regression coefficients for effects between latent variables without having to calculate factor scores.

Measurement Invariance

There are methods for both EFA and CFA for evaluating measurement invariance, which concerns whether scores from the operationalization of a construct have the same meaning under different conditions (Meade & Lautenschlager, 2004). These different conditions could involve consistency of measurement over populations, time of measurement, or methods of test administration (e.g., computer-administered vs. paper-and-paper format). Invariance over populations is related to the concept of construct bias, which implies that a test measures something different in one group (e.g., men) than in another (women). If not (i.e., there is no evidence for construct bias), then measurement is invariant over groups.

One method for testing invariance in EFA described by Thompson (2004, Chapter 8) involves best-fit rotation where the factor solution in one sample is specified as the target structure and the factor solution in a different sample is rotated to match the target structure as close as possible. (This assumes the same number of factors is retained in both samples.) The method derives a new matrix of pattern coefficients and structure coefficients for the non-target sample and also calculates factor correlations across the two samples. If the cross-sample factor correlations are low or there is no match of the pattern/structure coefficients across the groups, then measurement is not invariant. The method of CFA offers even more precise tests of measurement invariance. One rea-son is that it is possible in CFA to estimate mean contrasts on latent variables across ≥ 2 groups in a cross-sectional design or across ≥ 2 measurement occasions in longitudinal design for a given factor model. When means are analyzed in CFA, the model has a mean structure that represents factor mean differences and indicator intercepts for regression of the indicators on the factors in addition to the model's covariance structure. So specified, a CFA model analyzed across multiple groups can be tested for various levels of measurement invariance.

The most basic kind is configural invariance or equal form invariance. It is tested by specifying the same measurement model across the groups. In this model, both the number of factors and the factor-indicator correspondence are the same, but all parameters are freely estimated within each sample. If this model does not fit the data, then measurement invariance does not hold at any level. A stronger of form of invariance is construct-level

metric invariance, which means that the unstandardized pattern coefficients of each indicator are equal within the bounds of sampling error across the groups. If the construct-level metric invariance hypothesis is retained, then the researcher could conclude that the constructs are manifested the same way in each group. If some, but not all, of the pattern coefficients are equal, then there is evidence for partial measurement invariance. It also possible in CFA to test for invariance concerning factor variances and covariances, error variances and covariances, and indicator intercepts; see Kline (2010, chapters 9, 11) for examples.

Multilevel Factor Analysis

There are versions of both EFA and CFA for analyzing measurement models in hierarchical (nested) data sets where (a) scores are clustered into larger units and (b) scores within each level may not be independent. Suppose that data are collected from high school students who attend a total of 100 different schools. Students within each school are presumably affected by common characteristics that include the curriculum, teaching staff, school policies about discipline, student-to-teacher ratios, physical resources of the school, and so on. Scores from students who attend the same school may not be independent, and one aim of multilevel modeling or hierarchical linear modeling is to adjust the statistical estimates for the degree of score dependencies. Multilevel techniques are also used to estimate contextual effects of higher-order variables on scores of individuals in a hierarchical data set. An example in a multilevel factor analysis could be whether differences in student-teacher ratios between schools predicts the magnitude of covariation between indicators and factors within schools. Some SEM computer tools, including EQS, LISREL, and Mplus, support multilevel factor analysis, but the Mplus program is especially flexible in analyzing either exploratory or confirmatory measurement models in hierarchical data sets. See Heck and Thomas (2008) for an introduction to multilevel factor analysis.

Best Practices Recommended of EFA and CFA

Some best practices for factor analysis are briefly summarized: Report enough summary statistics, such as indicator correlations, standard deviations, and means, so that a reader could reproduce the results or test alternative models. Clearly spell out the rationale for indicator selection, measurement, model specification, data characteristics including score

reliabilities, and whether statistical assumptions are verified. Avoid applying factor analysis in samples that are just too small. Give specific details about decision points in the analysis. For example, state the data matrix analyzed, the method of factor extraction (EFA) or estimation (CFA), and the criteria for retaining a certain number of factors and selecting a particular rotation method (EFA). Report both the unstandardized solution and the standardized solution (with the appropriate standard errors) when analyzing a covariance matrix. If an initial model is respecified, inform readers about the theoretical or empirical justifications for these modifications. If no factor model is eventually retained, then explain what may be wrong with the theoretical foundations of the original model. Before using factor analysis, the researcher should thoroughly study the technique either in a course, professional workshop, or self-study, but he or she should be open to a process of continual learning. Along these lines, B. Thompson (2004) gives concise and clear introductions to EFA and CFA, Brown (2006) describes CFA for applied researchers, and Mulaik (2009) provides a comprehensive treatment of EFA for readers with strong quantitative backgrounds.

3.6. Practical Setting of Factor Analysis

3.6.1. Latent Variable Model for Factor Analysis

Consider the case of survey data, where we denote respondent *i's* answer to survey question j, $as\, x_{ij}$ $(i = 1,\ldots,n, j = 1,\ldots,k)$. Factor analysis hypothesises that x_{ij} is a combination of p unobserved factors, each written using the letter F. We have the following equation

$$X_{ij} = \lambda_{ji}F_{i1} + \lambda_{j2}F_{i2} + \cdots \lambda_{jp}F_{ip} + \delta_{ij} \tag{25}$$

where the k terms are factor loadings to be estimated, and δ_{ij} is the measurement error in X_{ij} or that part of X_{ij} that cannot be accounted for by the p underlying factors. It is possible to consider non-linear or multiplicative factor models but the simple linear, additive structure in equation above is by far the more widely used factor analysis model (Jackmann, 2002).

Factor Loadings

The k factor loadings are parameters to be estimated that tap how the unobserved factors account for the observed variables: the larger the values of k, the more a particular variable is said to "load" on the corresponding factor. Note that the k factor loadings vary across survey items, but not across individuals.

On the other hand, items vary in the way they are explained by the underlying factors, but the relationships between underlying factors and observed responses is constant across individuals (hence the absence of an i subscript indexing k). We note also that there are fewer underlying factors than there are variables $(p < k)$, consistent with the notion that like any statistical procedure, factor analysis is a device for "data reduction", taking a possibly rich though unwieldy set of survey responses and summarizing them with a simpler underlying structure.

Measurement Errors

The δ_{ij} terms for measurement errors simply reflect the idea that survey responses are not deterministically generated by the underlying factor structure. Like any statistical model, the factor structure is an approximation or a simplification that only captures so much of the survey responses under study.

Another way of thinking about measurement error is to imagine a respondent being asked to generate a response to a survey question on successive days: the observed responses would presumably fluctuate about a mean (tapped by the structural part of the factor analysis model), but the response on any given day will be a little above or below the average response. Conditional on the structure we postulate to underlie the survey responses, this random component of the survey response is tapped by δ_{ij} (Jackmann, 2002).

The Case of a Model for Multiple Responses

We can write an equation of the form of equation in section 3.5.6.2 above for each item being analyzed. Each equation expresses the corresponding survey response as a combination of "structure" and "noise": the underlying factors and measurement error,

respectively. Matrix notation allows the entire system of equations to be written quite efficiently: i.e., for each respondent;

$$X_i = \nu F_{i+} \delta_i \tag{26}$$

where $\boldsymbol{X_i}$ is a k by 1 vector of observed survey responses, $\boldsymbol{\nu}$ is a k by p matrix of factor loadings to be estimated, $\boldsymbol{F_i}$ is p by 1 vector of scores on the p underlying factors, and δ_{ij} is a k by 1 vector of measurement errors. In turn, we can lose the i subscript indexing individual respondents by stacking the above equation over respondents to yield $\boldsymbol{X} = \psi\nu' + \Delta$; where $\boldsymbol{X}$ is an n by k matrix of observed survey responses, $\boldsymbol{\psi}$ is an n by p matrix of scores on the underlying factors, $\boldsymbol{\nu}'$ is the transpose of the k by p matrix of factor loadings, and Δ is an n by k matrix of measurement errors.

Analysis of Covariance and Correlations

Since factor analysis usually works with the variances and co-variances of the observed X variables, it is sometimes referred to as "the analysis of covariance structures". Some hint of this is apparent in equation found in section 3.5.6.2, where the absence of an intercept term suggests that the means of the observed variables are either zero or of no direct interest. Indeed, this is typically the case in factor analysis, where the task is to learn about inter-relationships among variables rather than model the levels of each variable.

Moreover, it is generally not possible to estimate *both* the factor loadings and intercept terms (Jöreskog and Sörbom, 1989, Ch10). See also (Bollen, 1989, 306--311). Consequently, all the X variables and the unobserved n are presumed to have zero means, constraining any intercept term in equation in section 3.5.6.2 to zero. In addition, for the ordinal variables frequently encountered in surveys, the latent variable approach to generating a correlation matrix considers the variances of the latent variables to be equal 1, making all the co-variances between the latent variables interpretable as correlations. There exist constraints and identities that make the covariance structure representation of the factor analysis model tractable (Bartholomew et al., 2011).

3.6.2. Steps for Estimation in Factor analysis

Estimation via Principal Components or Eigen-Decomposition

Consider the case where $\boldsymbol{\gamma}$ is constrained to be an identity matrix, and so the model equation reduces to,

$$\sum = \nu\nu' + \varphi \tag{27}$$

The approach adopted in this thesis for the method for estimating the model parameters is principal components, exploiting the fact that a covariance matrix (i.e., a positive definite, square, symmetric matrix) can be decomposed as follows:

$$\sum = Z' \Gamma Z \tag{28}$$

Where $\boldsymbol{\Gamma}$ is a diagonal matrix containing the eigenvalues of $\sum$ in decreasing order $(Y_1 \geq Y_2 \geq \cdots, Y_k \geq 0)$ and $\boldsymbol{Z}$ is a $\boldsymbol{k}$ by $\boldsymbol{k}$ matrix of orthogonal eigenvectors. Each eigenvector can be usefully considered as a vector of coefficients that could be used for forming uncorrelated linear combinations of the X variables. For instance, using the jth eigenvector in this way produces a new variable $y_j = XZ_j$, which is the jth principal component of X (y_j is a n by 1 vector, X is an n by k matrix, and z_j is a k by 1 vector). Principal components have properties that make them especially useful for factor analysis.

The first principal component has the largest variance among all linear combinations of X. The second principal component as the second largest variance among linear combinations of X subject to the constraint that it is uncorrelated with the first principal component, and so on for subsequent principal components. Accordingly, each eigenvector of the correlation matrix is also a vector of principal components factor loadings.

While there are as many principal components as there are $\boldsymbol{X}$ variables, the idea behind factor analysis is to come up with a parsimonious representation of the structure underlying the $\boldsymbol{X}$ variables. In practice, then, only the first few principal components are retained, corresponding to a few factors. For any p factor model (with $p > k$), only the first p eigenvectors in $\boldsymbol{Z}$ are retained, and so the "full" k dimensional decomposition in equation is

not used; i.e., some of the variation in $\boldsymbol{X}$ is considered random, and is relegated to the $\boldsymbol{\varphi}$ matrix in the equation. The factor analysis model estimated by principal components is:

$$\sum = Z_{(p)}Z'_{(p)} + \varphi \tag{29}$$

Where $\boldsymbol{Z}(p)$ is the k by p matrix containing the first p eigenvectors of $\sum$. Another important property of factor analysis via principal components is that the model in equation is not unique.

Any rotation of $\boldsymbol{Z}$ that preserves its orthogonal structure fits the data just as well as the un-rotated solution in the equation. That is, the principal components factor loadings $\boldsymbol{Z}(p)$ can be multiplied by a *p* by *p* orthogonal matrix $\boldsymbol{G}$ to yield $\boldsymbol{Z^*_{(p)}} = \boldsymbol{Z_{(p)}G}$ and so;

$$\begin{aligned} \Sigma &= Z^*_{(p)}Z^{*\prime}_{(p)} + \varphi \\ &= (Z_{(p)}G)G'^{Z'_{(p)}} + \varphi \\ &= Z_{(p)}Z'_{(p)} + \varphi \end{aligned} \tag{30}$$

i.e., the factor loadings are identified only up to orthogonal rotations. The problem then becomes one of choosing among rotations that are optimal on other criteria. One popular choice is the *varimax* rotation (Kaiser, 1958), which produces factor loadings that have maximal variance, taking on values close to 1 and 0 in absolute value.

This helps ensure that the factors are reasonably distinct, with variables tending to load either quite strongly or quite weakly on any given factor. The researcher, in the analysis of his research data utilized the varimax rotation procedure.

The Maximum Likelihood Estimation Procedure

One other efficient approach is to estimate the parameters of the factor analysis model by the maximum likelihood Estimation (MLE). According to MLE, if $\mathrm{X_i}$' s are assumed to be iid Multivariate Normal i.e., $X_i \sim N(\mu, \Sigma)$, then the joint density of the data is;

$$F(X;\ \mu, \Sigma = (2\Pi)^{-\frac{kn}{2}} |\Sigma| \exp[-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)' \Sigma^{-1}(X_i - \mu)] \quad (31)$$

This provides the basis for maximum likelihood estimation: we can treat $\boldsymbol{\mu}$ as known (i.e., using the sample mean $\bar{X}$ which is the MLE of $\boldsymbol{\mu}$) and then embed the factor analysis model for $\boldsymbol{\Sigma}$ in the likelihood function. We work with the concentrated log likelihood (i.e., the log-likelihood that results from treating $\boldsymbol{\mu}$ as fixed at its sample estimate):

$$-\frac{n}{2} ln|2\Pi\Sigma| - \frac{n}{2} tr(S\Sigma^{-1}) \quad (32)$$

Where:

$$S = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$$

Anderson (2003, 14.3) considers properties of the maximum likelihood estimator for the orthogonal factor model $\widehat{\Sigma} = \widehat{\Lambda}\,\widehat{\Lambda}' + \varphi$ also Mardia, Kent and Bibby (1979) and Lawley and Maxwell (1971) share his views. As a practical matter, and at least for the orthogonal factor model, it is easier to *minimize* the following function with respect to Σ,

$$F(\Lambda, \varphi;\ S) = tr(\Sigma^{-1}S) - ln|\Sigma^{-1}S| - K \quad (33)$$

i.e., minimizing this equation wrt $\boldsymbol{\Sigma}$ yields the same result as maximizing the log-likelihood in the previous equation. Computational strategies for carrying out this optimization are discussed in (Mardia, Kent and Bibby,1979, 264-266), summarizing the pioneering work by (Jöreskog 1967).

The Likelihood Ratio Test Statistic

Consider the case where the factor analysis model fits the data perfectly: that is, the observed covariance matrix S is perfectly recovered by Σ. In this case the log-likelihood reduces to

$$ln(\mathcal{L}_o) = -\frac{n}{2} ln|2\pi S| - \frac{nk}{2}, \quad (34)$$

Since in this case $S = \widehat{\Sigma}$ and the trace of $S\hat{}\ (\widehat{\Sigma}^{-1}) = I_k$ is *k*.

Comparing the two log-likelihoods in above equations, we construct a likelihood ratio test statistic follows:

$$q = \frac{\mathcal{L}}{\mathcal{L}_o} \tag{35}$$

$$\ln(q) = \ln(\mathcal{L}) - \ln(\mathcal{L}_{o)}$$

and then

$$\text{-}2\ln(q) \sim \chi_v^2$$

Where the degrees of freedom parameters *v* is the number of unique elements in the covariance matrix **S** minus the number of parameters estimated, or

$$\begin{aligned} v &= [\tfrac{1}{2}(k+1)k] - [k + kp - \tfrac{1}{2}p(p-1)] \\ &= \frac{1}{2}[k^2 + k - 2k - 2pk + p^2 - p] \\ &= \tfrac{1}{2}[(k-p)^2 - (k+p) \end{aligned} \tag{36}$$

where k is the number of $\boldsymbol{x}$ variables, and p is the number of factors being estimated. If the difference in the unconstrained and constrained likelihoods is large, then the test statistic will be large, and, if sufficiently large, we will reject the constrained model in favour of a less constrained model. One of the key features of the representation of the optimization problem underlying maximum likelihood factor analysis given in the theory is that it is leads directly to the test statistic given above. That is, the test statistic;

$$-2lnq = nF(\hat{\Lambda}, \varphi) \tag{37}$$

where *F* is as defined in earlier equations, has an asymptotic χ_v^2 distribution, where the degrees of freedom *v* is been defined above (Rice, 2010).

Estimation by Weighted Least Squares

According to (Bollen, 1989, 425), a weighted least squares estimator finds estimates $\hat{\sum}$ that minimize the criterion

$$= [vec(\boldsymbol{S}) - vec(\hat{\sum})]' \boldsymbol{\varphi}^{-1}[vec(\boldsymbol{S}) - vec(\hat{\sum})], \quad (38)$$

Where $\boldsymbol{\varphi}$ is a matrix of weights, and the vec operator turns the lower triangle of its matrix argument into a vector. If $\boldsymbol{S}$ is a k by k matrix then $vec(\boldsymbol{S})$ is a vector of length $k(k+1)/2$, containing the unique elements of $\boldsymbol{S}$ (k diagonal elements, plus $k(k-1)/2$ unique off-diagonal elements). Accordingly, $\boldsymbol{\varphi}$ is a $k(k+1)/2$ by $k(k+1)/2$ matrix. If the elements of $\boldsymbol{\varphi}$ contain consistent estimates of the variances and covariances of the $vec\,(\boldsymbol{S})$, then the Browne (1984) asymptotically-best distribution-free WLS estimator results from choosing $\hat{\sum}$ to minimize the criterion in equation above. This ADF-WLS estimator is especially useful when working with ordinal survey responses, where non-normality in the "normal scores" is almost guaranteed.

The parameter $\boldsymbol{\varphi}$ is usually estimated in the course of generating $\boldsymbol{S}$ from the normal score representation of the ordinal responses. Each diagonal element of $\boldsymbol{\varphi}$ is an estimate of the variance of the corresponding sample correlation in $vec(\boldsymbol{S})$, while each off-diagonal elements is an estimate of the covariance in pairs of sample correlations. The off-diagonal quantities turn out to be functions of the cross-kurtoses in the raw data; if the raw data were distributed iid multivariate normal, then a consistent estimate of $\boldsymbol{\varphi}$ would be an identity matrix. The ADF-WLS estimator is distribution free in the sense that the raw data can be of almost any distribution, but with an appropriate $\boldsymbol{\varphi}$ matrix we can obtain estimates with properties such as consistency and asymptotic normality. This makes asymptotically-valid hypothesis testing and inference-making possible even when the raw data are non-normal.

3.6.3. Understanding the Effective Number of Free Parameters

We note from the orthogonal factor analysis that the model implies by (Lawley and Maxwell, 1971),

$$\Sigma - \varphi = \Lambda\Lambda'. \tag{39}$$

Suppose now that each variable in X is rescaled so that the residual measurement error variances (the diagonal elements of φ) are all one. This means that $\Sigma^* = \Lambda^* \Lambda^{*\prime} + \mathrm{I}$ which follows from transforming the original model

$$\varphi^{-\frac{1}{2}} \Sigma \varphi^{-\frac{1}{2}} = \varphi^{-\frac{1}{2}} \Lambda\Lambda' \varphi^{-\frac{1}{2}} + \varphi^{-\frac{1}{2}} \varphi \varphi^{-\frac{1}{2}} \tag{40}$$

$$\text{and hence } \Sigma^* = \varphi^{-\frac{1}{2}} \Sigma \varphi^{-\frac{1}{2}}.$$

Furthermore, this transformation of the model implies that Σ^* - φ is transformed to become

$$\varphi^{-\frac{1}{2}}(\Sigma - \varphi)\varphi^{-\frac{1}{2}} = \varphi^{-\frac{1}{2}} \Sigma \phi^{-\frac{1}{2}} - \phi^{-\frac{1}{2}} \varphi \varphi^{-\frac{1}{2}} \tag{41}$$

$$= \Sigma^* - I_k$$

which is symmetric and has rank p. Accordingly, we can decompose $\Sigma^* - I$ into

$$\Omega Y \Omega' \tag{42}$$

where Y is a diagonal matrix of order p and Ω is a k by p matrix such that

$$\Omega'\Omega = I_p. \tag{43}$$

The elements of Y contain the p non-zero eigenvalues of $\Sigma^* - Ik$ and the columns of Ω are the corresponding eigenvectors. This decomposition implies a unique solution for Λ.

$$\Lambda = \varphi^{\frac{1}{2}}\Omega Y^{\frac{1}{2}} \tag{44}$$

This holds since

$$\Lambda\Lambda' = \varphi^{\frac{1}{2}}\Omega Y^{\frac{1}{2}}Y^{\frac{1}{2}}\Omega'\varphi^{\frac{1}{2}} \tag{45}$$

$$= \varphi^{\frac{1}{2}}\Omega\, Y\Omega'\varphi^{\frac{1}{2}}$$

$$= \varphi^{\frac{1}{2}}(\Sigma^{*} - \mathrm{I}_{k})\varphi^{\frac{1}{2}}$$

$$= \Sigma^{*} - \varphi$$

Which is required by the model. Further, intuitively, we determine that

$$\Lambda'\varphi^{-1}\Lambda = Y^{\frac{1}{2}}\,\Omega'\varphi^{\frac{1}{2}}\varphi^{-1}\varphi^{\frac{1}{2}}\Omega Y^{\frac{1}{2}} \tag{46}$$

$$= \mathrm{Y}^{\frac{1}{2}}\,\Omega'\Omega\mathrm{Y}^{\frac{1}{2}}$$

$$= Y$$

This is so because

$$\Omega'\Omega = \mathrm{I}_{p} \tag{47}$$

But Y is a diagonal matrix of order P, which is a condition imposing constraints on Λ and ϕ. This means there are kp free parameters in Λ, and k free parameters in $\boldsymbol{\varphi}$, which is a requirement that $\Lambda'\varphi^{-1}\Lambda$ be diagonal imposes $\frac{1}{2}p(p-1)$ on the model parameters. In the absence of these constraints on the p *by* p symmetric matrix $\Lambda'\varphi^{-1}\Lambda$, there would be $\frac{1}{2}p(p+1)$ free parameters.

The orthogonal, unit variance p factor analysis model has:

- Kp factor loadings to estimate (each of k variables loading onto all p factors) in Λ;

- K measurement error variances to estimate (no error variances in $\boldsymbol{\varphi}$;
- Less $\frac{1}{2}p(p-1)$ parameters imposed by the constraint that $\Lambda'\varphi^{-1}\Lambda$ must be diagonal, for a total of $k + kp - \frac{1}{2}p(p-1)$ free parameters.

3.6.4. Goodness of Fit procedure

As noted by (Jackman, 2005), when dealing with orthogonal factors, the relative magnitudes of the eigenvalues determine the proportion of the variance explained. This holds for any orthogonal rotation of a principal components solution such as the common varimax rotation. There is also a way to do residual analysis with the factor analysis model. Note the model we fit was $\hat{\sum} = \widehat{\Lambda}'\widehat{\Lambda} + \hat{\varphi}$ with a perfect fit being when $\hat{\sum} = \boldsymbol{S}$ (the sample covariance or correlation matrix).The simplest summary measure of goodness-of-fit involves simply comparing $\hat{\sum}$ with S.

One should always inspect this "residual matrix" (S - $\hat{\sum}$) for large elements which suggest model inadequacy; note that the matrix will be symmetric and thus 10 have only $k(k-1)/2$ unique elements. Various summary measures have been proposed. One popular candidate is root mean-square residual (RMR),

$$= \left[\mathbf{2}\sum_{i=1}^{k}\sum_{j=1}^{i}\frac{(S_{ij-}\sigma_{ij})^2}{k(k+1)}\right]^{\frac{1}{2}} \tag{48}$$

i.e. the square-root of the mean of the squared elements of the residual matrix.

3.6.5. Determination of Eigen Values and Eigen vectors

Matrix A Acts by Stretching the Vector x, not Changing its Direction, so x is an Eigenvector of A

In many contexts, a vector can be defined as a list of real numbers (called elements), written vertically with brackets around the entire list. Two vectors are said to be scalar multiples of each other (also called parallel or collinear) if they have the same number of elements, and if every element of one vector is obtained by multiplying each corresponding element in the other vector by the same number (known as a scaling factor, or a scalar).

For example, take the case where vectors are scalar multiples of each other, because each element of v is −20 times the corresponding element of u.

A vector with three elements, like u or v, may represent a point in three-dimensional space, relative to some Cartesian coordinate system. It helps to think of such a vector as the tip of an arrow whose tail is at the origin of the coordinate system. In this case, the condition "u is parallel to v" means that the two arrows lie on the same straight line, and may differ only in length and direction along that line.

If we multiply any square matrix A with n rows and n columns by such a vector v, the result will be another vector $w = Av$, also with n rows and one column. That is,

(49)

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_{\textit{is mapped to}} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \dots & A_{n,n} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Where, for each index i,

$$w_i = A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,n}v_n = \sum_{j=1}^{n} A_{i,j}v_j$$

In general, if v_j are not all zeros, the vectors v and Av will not be parallel. When they are parallel (that is, when there is some real number λ such that

$$Av = \lambda v \tag{50}$$

we say that v is an eigenvector of A. In that case, the scale factor λ is said to be the eigenvalue corresponding to that eigenvector. In particular, multiplication by a 3×3 matrix A may change both the direction and the magnitude of an arrow v in three-dimensional space. However, if v is an eigenvector of A with eigenvalue λ, the operation

may only change its length, and either keep its direction or flip it (make the arrow point in the exact opposite direction).

Specifically, the length of the arrow will increase if $|\lambda| > 1$, remain the same if $|\lambda| = 1$, and decrease it if $|\lambda| < 1$. Moreover, the direction will be precisely the same if $\lambda > 0$, and flipped if $\lambda < 0$. If $\lambda = 0$, then the length of the arrow becomes zero.

Statistical Analysis and Output

The initial data analysis output were presented as $mean \pm SD$. Only participants with complete data for all the variables of interest were considered in the present study.

Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data.

Factor analysis is related to principal component analysis (PCA), but the two are not identical (Bartholomew et al., 2008). Because PCA performs a variance-maximizing rotation of the variable space, it takes into account all variability in the variables. In contrast, factor analysis estimates how much of the variability is due to common factors ("communality"). The two methods become essentially equivalent if the error terms in the factor analysis model (the variability not explained by common factors, see below) can be assumed that all have the same variance.

Several statistical methods can be used to identify patterns of clustering in cardiovascular diseases such as DM and hypertension. One such important and useful technique is factor analysis – a multivariate technique. Indeed, Factor analysis is a statistical method used to describe variability among observed variables in terms of a potentially lower number of unobserved variables called factors. This includes two types for implementation: Exploratory factor analysis and confirmatory factor analysis. In other words, it is possible, for example, that variations in three or four observed variables mainly reflect the variations in a single unobserved variable, or in a reduced number of unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables.

The observed variables are modeled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset. The definition of MS is controversial because of the number and the pre-requisite criteria. Indeed, the World Health Organization (WHO) and the European Group for Insulin Resistance (EGIR) consider Insulin resistance (Hyperglycemia) as the central role for MS, whereas the US (ATP III NCEP) and the International Federation for Diabetes (IFD) considered abdominal obesity as the pre-requisite and central role of MS. The consensus from IFD recommended the Europid cut-offs of waist circumference (WC >94 cm for men and >80 cm for women) for sub-Saharan Africans before the lack of valid data.

However, WC>102 cm for men and WC>88 cm for women are used to define MS in the US (Chad et al., 2013). As the lipid profile (Total Choresterol and Tryglycerides) is often within the normal range, and both lower and higher levels of high density lipoprotein-Cholesterol (HDL-C) are associated with Atherosclerosis (Stroke, Mycardial Infarction). Thus, thanks to African valid data defining abdominal obesity; MS is now defined for sub-Saharan Africans using WC>94 cm for both men and women. Moreover, abdominal obesity is the central role of MS in western and central African type 2 diabetics with similar WC level in comparison with non-classified diabetic central Africans. Metabolic syndrome defined by international cut-off values are limited to detection at high metabolic risks among Central Africans in comparison with metabolic syndromes defined by ethnic-specific definitions.

There is a significant U-shaped relationship between arthrosclerosis complications, insulin resistance and HDL-Cholesterol stratification. In general, Bantu Africans with and without adulthood diabetes mellitus (DM) and/or MS may be ethnically characterized by a particular clustering of components of MS. For that reason, the objective of this study was to provide a step-by-step description of the application of factor analysis and interpretations of the results based on anthropometric parameters, blood pressure and plasma glucose in the general population, men, women, rural and urban inhabitants and different types of DM.

Data were presented as mean ± SD. Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations

research, and other applied sciences that deal with large quantities of data. Factor analysis is based on the following statistical model and definitions. Suppose we have a set of p observable random variables. Suppose further that we have a set of p observable random variables, $x_1, \ldots, x_p$ with means $\mu_1, \ldots, \mu_p$.

Note that for any Orthogonal Matrix Q, if we set L = LQ and F = Q^TF, the criteria for being factors and factor loadings still hold. Note that for any Orthogonal Matrix Q, if we set L=LQ and F=Q^TF, the criteria for being factors and factor loadings still hold. Hence a set of factors and factor loadings is identical only up to orthogonal transformations. Common factor analysis, also called principal factor analysis (PFA) or principal axis factoring (PAF), seeks the least number of factors which can account for the common variance (correlation) of a set of variables.

Analogous to Pearson's r, the squared factor loading is the percent of variance in that indicator variable explained by the factor. To get the percent of variance in all the variables accounted for by each factor, the sum of the squared factor loadings for that factor (column) was added and divided by the number of variables. This is the same as dividing the factor's Eigenvalue by the number of variables. The Eigenvalue for a given factor measured the variance in all the variables which is accounted for by that factor. Eigenvalues measure the amount of variation in the total sample accounted for by each factor. Extraction sums of squared loadings were performed.

Factor scores were the scores of each case (row) on each factor (column). To compute the factor score for a given case for a given factor, the case's standardized score was taken on each variable, multiplied by the corresponding factor loading of the variable for the given factor; and these products were summed. For determining the number of factors, the Kaiser criterion was used.

The Kaiser rule is to drop all components with Eigenvalues less than 1.0. The Cattell scree test plotted the components as the $X - axis$ and the corresponding eigenvalues as the $Y - axis$. As one moves to the right, toward later components, the eigenvalues drop. When the drop ceases and the curve makes an elbow toward less steep decline, Cattell's scree test

says to drop all further components after the one starting the elbow. Varimax Rotation served to make the output more understandable and facilitated the interpretation of factors.

This is an orthogonal rotation of the factor axes to maximize the variance of the squared loadings of a factor (column) on all the variables (rows) in a factor matrix, which has the effect of differentiating the original variables by the extracted factor. Further, we used oblique Promax rotation as an additional alternative to varimax rotation for suitable clustering characteristics. A $P - value < 0.05$ was considered as statistically significant. All analyses were performed using the Statistical Package for Social Sciences (SPSS) for windows version 21.0 (SPSS Inc, Chicago, Il, USA).

3.7. Conclusion

This topic focused on the methodology for this research. We chose to put more emphasis on the data collection strategies followed by the analysis which have been distributed on the basis of the requirements of the Degree of Doctor of Philosophy. While the data collection has been detailed in this chapter, the next chapter has focused on the nature and type of data for this study. The chapter has seen the researcher create different data collection approaches for different data variables and further, an outline of the sources of the data collection method. The researcher has explained in detail the method used for each and every specific variable. For example, the method used to collect waist circumference has been shown to be quite different from that used to collect weight, etc.

In addition, the researcher has explained the origins of the different variables used in this research. Some approaches of data analysis have been explained in this chapter. The analysis was varied where different ingredients under factor analysis have been employed. The researcher has included the formulas required for this study. A number of formulas have assumed that the reader has no knowledge or only basic know-how. Conclusions have been drawn based on the outcome of the analysis.

Chapter 4

4. The Data

4.1. Introduction

This chapter discusses the survey procedure and the data used for this study. These surveys captured medical data on individuals who either suffered from factor(s) of Metabolic Syndrome or not. In the sections that follow, the survey design was discussed after which each survey was considered independently. The discussion includes the data collection procedure; the area of study and the strategies employed to collect information and as well as estimation of sample size for the survey. The final section in this chapter covers data coding and capturing.

The researcher started by details of the type of data collected. It will be observed that the data used were collected on some of the following variables which have been explained elsewhere; Diastolic Blood Pressure, Systolic Blood Pressure, Waist Circumference, Body Mass Index, Fasting glucose, Post Load Glucose, etc. These were measurable variables and so some of the data were basically quantitative but continuous. This is due to the fact that all the variables included fractions (decimal points). Below, the researcher gives definitions of different types of variables with particular reference and emphasis on the variables which were utilized for this study.

4.2. Types of Data

Medical research often involves collecting information about biological parameters, such as measurements of blood pressure or pulse rate, etc. on a group of individuals. Many biological characteristics resulting measurements vary from person to person and are referred to as variables.

The measurement unit is 'millimeters of mercury' (mmHg) and is written as two numbers. For example, if your reading is 120/80mmHg, your blood pressure is '120 over 80'.Every blood pressure reading consists of two numbers or levels. They are shown as one number on top of the other. The first (or top) number is your systolic blood pressure. It is the highest level your blood pressure reaches when your heart beats. The second (or bottom) number is your diastolic blood pressure. It is the lowest level your blood pressure reaches as your heart relaxes between beats.

Similarly, waist circumference, Body Mass Index, Post Load Glucose and Fasting Glucose are measured as continuous anthropometric variables for which the first two variables use centimeters as the unit of measurement. The following procedure and explanation states how glucose is measured.

Insulin hormones from the pancreas regulate the concentration of blood sugar, or glucose, in the body. When blood glucose rises, insulin is released into the blood to utilize the glucose for energy. People with diabetes have difficulty producing insulin; thus, their blood glucose levels remain high. High glucose levels in the body can negatively affect a number of internal organs. Fasting blood sugar is the amount of glucose in your body after you've not eaten for at least 8 hours. Measuring your fasting blood sugar is one way to check for diabetes. This is measured in mg per dL. A meter reading of 100 mg per dL or fewer means that your blood glucose levels are normal. A measurement of 100 mg per dL means that there are 100 milligrams of glucose for every deciliter of blood in your system. Imbalances of blood sugar are common among patients of diabetes mellitus. Diabetes indicates persistently high blood sugar that may cause damage to various organs like the kidney, heart, small arteries and the eyes (retina). Diabetics are also prone to sudden drops in blood sugar called hypoglycemia. To monitor these fluctuations, blood sugar measurements are vital to diabetic individuals.

4.3. Area of the Survey (Kinshasa Hinterland)

The Geography of Kinshasa Hinterland

Figure 7 The geographic characteristics of Kinshasa Hinterland.

By definition, "Hinterland" is a German word meaning "the land behind" (a city, a port, or similar). The term was first used in English in 1888 by George Chisholm in his work Handbook of Commercial Geography. Kinshasa is a Metropolitan Municipality City in Congo located in the South Western region of the country (Figure 7). The reader must be informed that Kinshasa is the capital city of the Democratic Republic of Congo. It is situated on the border with the republic of Congo Brazaville. Kinshasa is surrounded by the cities of Ilebo on the east, Kikwit on the south and Matadi on the south west. However, as a country Congo is bordered by nine African countries namely; Central African Republic, Sudan, Uganda, Rwanda, Burundi, Tanzania, Zambia, Angola, and Republic of the Congo.

Kinshasa, formerly (until 1966) Léopoldville, the largest city and capital of the Democratic Republic of Congo. It lies about 320 miles (515 km) from the Atlantic Ocean on the south bank of the Congo

River. One of the largest cities of sub-Saharan Africa, it is a special political unit equivalent to a Congolese region, with its own governor. The city's inhabitants are popularly known as Kinois. Kinshasa is not only the capital but also the center of the dynamic and contradictory influences that have shaped the country's character in modern Africa. The only city, not clearly identified with any particular region of the country, was until 1997 the seat of a long-lasting Zairean military government base, on the one hand, on the strength of the armed forces and, on the other, on a technique of political and social compromise that until its later years, gained the rather grudging collaboration of most of the citizens. Caught between spectacular wealth and massive poverty, most Kinois must spend a considerable amount of their time scrambling for necessities that are in erratic supply. Nevertheless, they have found the means to make Kinshasa a source of distinctive influence in intellectual and popular culture felt throughout Africa. The most heavily inhabited area of Kinshasa covers 58 square miles (about 150 square km).

The total area subject to city government, much of it sparsely populated, is 3,848 square miles (9,965 square km). The estimated Population (2005 est.) was 5,717,000. Kinshasa spreads out southward from the shoreline of the Congo River at Malebo Pool, a widening of the river. The plain on which the city lies varies mostly between 918 and 1,148 feet (280 and 350 metres) above sea level and is partly encircled by higher ground. The surrounding countryside is heavily farmed savanna and gallery forest; the chief crops being cassava, sugarcane, oil palms, plantains, corn (maize), peanuts (groundnuts), and beans.

The climate is hot year-round, with a dry season from May to September and a rainy season from October to May. The mean annual rainfall is slightly more than 60 inches (1,520 mm). Violent rainstorms occur frequently but seldom last more than a few hours. The hottest month is April, with mean daily maximum and minimum temperatures of 89 °F (32 °C) and 71 °F (22 °C), respectively. The corresponding figures for July, the coolest month, are 81 °F (27 °C) and 64 °F (18 °C). Higher suburbs are somewhat cooler than the central city.

The built-up area of Kinshasa is divided into industrial, residential, and commercial zones. Along the western edge of the central city, an industrial zone (before 1966 called Léo-Ouest) flourishes near the site of the first depot established by the British-American explorer Sir Henry Morton Stanley. To its east lies the riverside residential and administrative district of Gombe, which houses most of the European population and the Congolese elite; the central government buildings and

the embassy district are located there. The eastern sector (known before 1966 as Léo-Est), of which the wide Boulevard du 30-Juin forms the main artery, is a major commercial area. The waterfront, along Kinshasa's northern edge, is lined with quays and large warehouses. Ndolo, east of Gombe, comprises a complex of port facilities and industrial plants. The poorer areas extend southward on the east and west of Kinshasa. Among Kinshasa's satellite cities, Ndjili, to the southeast, has become a residential area, while Kimpoko upstream has been developed as an outer port. During the 1970s wealthy businessmen and politicians built mansions, often of spectacular opulence, in Binza, an area in the western hills overlooking the city.

The population of Kinshasa grew slowly at first (from 5,000 people in 1889 to 23,000 in 1923) but increased rapidly after 1940. The population doubled from 1950 and this trend continued thereafter about every five years and by the beginning of the 21st century it approached 5,000,000, a majority of whom lived in the squatting zones. Much of the population growth has been the result of Congolese migration and government expansion, but widening of the city's boundaries has contributed to the increase. Kinshasa has a young population.

More than half the people are under 22 years of age, and only a tiny proportion of the population is over 50. Migration of people from the rural areas intensified greatly after independence as colonial restrictions were relaxed. Political troubles and the economic decline of rural areas and their lack of amenities and opportunities, as well as the attractions of the city, have contributed to this rural exodus. In its early years, the city received immigrants from western Africa and the various neighboring countries of Central Africa; since independence, however, most new inhabitants have come from within Congo, especially the nearby regions of Bandundu to the west and Bas-Congo (Lower Congo) to the south and east.

4.4. Health Conditions of the People of Kinshasa Pertaining to Non-Communicable Diseases

A number of non-communicable diseases have been reported by different medical researchers. It was pains-taking to get information about a few of the non-communicable diseases engulfing the area of Kinshasa. Research reports indicate that due to change of lifestyles, the majority of inhabitants have adopted the lifestyle of changing their dietary habits, physical inactivity, (Hilawe et al., 2012). A Western lifestyle and changing eating habits from African nutritious foods to English foods has created a healthy problem of enormous nature. The researcher outlines a few of the observed diseases from research literature.

4.5. Depression

Depression was common after stroke with an occurrence of 53.6%. These results highlight the need to investigate, diagnose and treat PSD, which is a risk factor for morbidity and mortality after stroke. The World Health Organization (WHO) defines stroke as the rapid development of clinical signs of local or global cerebral dysfunction with symptoms lasting at least 24 hours or leading to death with no apparent cause other than the vascular origin. There are two types of stroke, depending on the etiology: hemorrhagic stroke and ischemic stroke. Stroke now constitutes a public health problem throughout the world and is the second cause of death worldwide. In the Western countries, stroke is the third largest cause of death after cancer and myocardial infarction and remains the leading cause of dependency. Africa is currently experiencing an epidemiological transition. The World Health Organization has highlighted the emergence of non-communicable chronic diseases, including stroke, in developing countries. As a cause of death, stroke ranks first in Africa, above infectious diseases (Mpembi et al., 2013). Hypertension.

The prevalence of prehypertension and hypertension, their determinants and associated cardiovascular risk factors in Congolese urban dwellers are factors that have been observed by medical researchers. A study carried out on Congolese urban men and women of varying ages showed that: prehypertension was observed in 30.3% of subjects, 34.9% of men and 26.7% of women ($P = 0.0045$). The prevalence of hypertension amounted to 30.9% with no difference between genders. Participants with prehypertension had average age, BMI and waist circumference intermediate between those with normal BP and hypertensive subjects.

Their glucose and lipids levels were similar to those of normotensives. The prevalence of prehypertension amounted to 33% at age 20 - 29 years and decreased to 16.7% at ≥60 years whereas the prevalence of hypertension increased from 11.2% to 71.4%. The rates of diabetes mellitus were similar across blood pressure categories whilst prevalence of overweight/obesity, abdominal adiposity, dyslipidemia and metabolic syndrome significantly increased (P = 0.05 or less). Among participants with prehypertension, 73% had two or more additional cardiovascular risk factors. In the logistic model the probability of prehypertension was higher in men (OR: 1.429; 95% CI: 1.099 - 1.857) and participants with overweight/obesity (OR: 1.666; 1.146 - 2.422), lower in participants aged ≥ 55 years (Bayauli et al., 2014).

4.6. Diabetes

Increasing urbanization and the accompanying changes in lifestyle are leading to a burgeoning epidemic of chronic non-communicable diseases in sub-Saharan Africa. At the same time, the prevalence of many acute communicable diseases is decreasing. In consequence, the inhabitants of sub-Saharan Africa are generally living longer and this increasing longevity will result in a rise in the future incidence of non-communicable diseases in the region. Diabetes mellitus is one of the most prominent non-communicable diseases that are undermining the health of the people in sub-Saharan Africa and placing additional burdens on health systems that are often already strained. In 2011, 14.7 million adults in the African Region of the World Health Organization (WHO) were estimated (Hilawe et al., 2013).

Longo-Mbenza et al., 2010) confirm by claiming that out of 3,156 person-years of median follow up, there were 93 type 2 diabetes cases (11.5%), corresponding to an incidence of 29 (95% CI 15—43) per 1,000 person-years. The independent predictors of incident type 2 diabetes were age ≥ 45 years, physical inactivity, non-diabetic hyperglycaemia and metabolic syndrome regardless of the criteria used. In conclusion, urgent prevention strategies are needed to curb the type 2 diabetes epidemic in Africa. This group of researchers reaffirm their claim by adding, a high degree of prevalence of Cardio-metabolic Syndrome in the Congolese urban communities (Longo-Mbenza et al., 2014).

On a more serious note, further studies on the prevalence rates and cardio-metabolic determinants of diabetes mellitus and pre-diabetes with projected coronary heart disease at the bank site of Brazzaville revealed that out of the employees, 16% and 21.4% had DM and pre-diabetes, respectively. The rate of T2DM among diabetics was estimated 90%. Aging, high total cholesterol, high LDL-cholesterol, high conicity index and longer urban residence after migration were significantly associated with pre-diabetes. Physical inactivity, smoking, excessive alcohol intake, abdominal obesity, female gender, low HDL-C, hypertension, CHD, projected high 10-year total CHD risk, age ≥ 55 years, urban residence, Southern area residence, high socioeconomic status, non-married status, MetS/NCEP, MetS/IDF for Europe and MetS/IDF for Africa were significantly associated with T2DM. MetS/IDF for Africa was the only independent determinant of T2DM (Thierry et al., 2014).

4.7. Economy, Commerce and Industry

Commerce and Industry

Kinshasa is the most important consumer centre of the republic and the core of its industrial and commercial activity. The city serves as the headquarters of major public corporations and of privately owned industrial and commercial companies. It dominates the financial and commercial life of the republic and houses the head offices of principal banks. Among Kinshasa's main industries are food processing and those producing consumer goods (e.g., beer, textiles, and footwear), generally for domestic markets. Construction and various service industries also contribute to the city's economy. However, the political turmoil that has gripped the country since the downfall of the Zairian regime in 1997 has been debilitating the city's economic activities. Kinshasa, the capital city of the Democratic Republic of Congo, is one of Africa's fastest-growing cities. Some 12 million people live there; in Africa only Lagos and Cairo are larger. It is a hard place to live, says Mr Mbalane, a resident for 27 years. The nightlife may be vibrant, but the streets are filthy. And with Congo's president, Joseph Kabila, up against term limits next year, with no clear successor, no one takes peace for granted.

Congo's horrific civil war, fought mostly in the east, far from Kinshasa is winding down. It split the country into warring territories and claimed somewhere between 800,000 and 5 million lives, depending on which estimate you distrust less. Over the past decade, Congo has reunited, more or less ethnic militias that once controlled vast swathes of territories no longer do so. Peace has come economic growth, which is estimated to have averaged more than 7% a year since 2009 and 9% in 2014. Inflation is low; creating a stable exchange rate. Aid has surged into Kinshasa even as a mining boom has filled public coffers. The capital has had a facelift: the Boulevard du 30 Juin, its main thoroughfare, is now a modern highway; the airport has a gleaming new terminal from which most of the rapacious officials who used to fleece travelers have been banished.

The gains have been unequally distributed, however. In Gombe, the central, Belgian-planned district where expats and the Kinois (Kinshasan) elite live, new apartments are sprouting up. Developers are spilling outwards: one project, called the "Cité du Fleuve" (River City), is a self-contained block of plush flats built on reclaimed land, selling for as much as $900,000 each. Posh nightclubs are full on weekends; hip restaurants serve steaming plates of *cossa cossa* (giant river prawns). Even the music scene, which had withered, is recovering as Congolese musicians flock

back from Europe to play for local yuppies. "It was all in Paris, but now, Kinshasa is the place to be," says Reddy Amisi, a musician who lives in France but returns to Kinshasa every few months.

Outside this bubble, however, life is different. Most Kinois live in crowded slums with neither electricity nor clean water. When the river is high, their homes flood. Over a third of the population is younger than 15; of the adults, fewer than 10% have salaried jobs. There are, however, some signs of progress. The streets are full of motorbike taxis, which were almost unheard of a few years ago. (They are known as *wewa*, which means "you" in the language of many drivers and is what Kinois shout to hail one.) At the Marché de la Liberté, a sprawling market, traders do a roaring business in food, clothes and mobile phones. But the cost of living is high, and widespread corruption makes it hard for honest folk to get ahead. "I sometimes make a lot of money, but always it is taken by the police," says José Kalenda, a mobile-phone trader. This divide is what makes Congo unstable.

The state is perilously fragile. It all but collapsed under Mobutu Sese Seko, a flamboyant despot who allowed officials to steal so that he didn't have to pay them. Since 2001 the country has been run by Joseph Kabila, the son of the man who overthrew Mobutu in 1997 with a lot of help from Rwanda, Congo's neighbour. Kabila senior ruled atrociously until his assassination in 2001. His son quickly assumed the presidency and was elected to the job in 2006. The constitution, adopted in 2006, says he must stand down next year. But many doubt he will. Officials have suggested that elections due to happen in November 2016 will have to be delayed, perhaps for years. They talk of the need to conduct a census first—an impossible task in a continent-sized country with hardly any decent roads.

4.8. Kinshasa: The key of the Democratic Republic of Congo

Few in the capital support the president. He is seen as an outsider. Having been educated in Tanzania, he struggles with both French and Lingala, the two local languages. Slum-dwellers see money flowing into the capital but never reaching them. If Mr Kabila refuses to step down when his term is over, protests will surely erupt. In January, during demonstrations against electoral rule-changes, Congolese troops killed about 40 people. Many fear a larger revolt next year. "Kinshasa is the key" to Congo, says an expatriate businessman.

Congo's problems are a grander, more dangerous version of what is happening in neighbouring countries. In October, in next-door Congo-Brazzaville, troops fired on protesters who objected to

President Denis Sassou Nguesso's plan to extend his three-decade rule (the protesters' slogan was "Sassoufit", a play in the French for "That's enough"). In Burundi several hundred people have been killed since President Pierre Nkurunziza said he would run for a third term. If a similar dynamic plays out in Kinshasa, the result could be far more destructive. The mining boom has stalled, thanks to the drop in commodity prices. If Mr Kabila ever fails to pay the army properly, all bets are off. For the moment, however, most Kinois prefer not to think about such things. James Peter, who runs a stall at the Marché de la Liberté selling bathroom fittings, says that the problem with politics is that when elections are approaching, nobody wants to buy anything. He doesn't have time for debates about who will succeed Mr Kabila; he has three children to send to school.

The business climate in Congo "is disgusting", says an adviser to the government in Kinshasa. Any casual visitor has probably noticed. Traffic police stop cars for no reason, force their way in and refuse to leave until paid off. Tax agents arrive at company offices with seven and eight-figure demands that, of course can be negotiated down.

A countable number wonder this central African nation's biggest business digging in the dirt to extract precious minerals—is so dirty. An expert panel led by Kofi Annan, a former UN secretary-general, looked at five deals struck between 2010 and 2012, and compared the sums for which government-owned mines were sold with independent assessments of their value. It found a gap of $1.36 billion, double the state's annual budget for health and education. And these deals are just a small subset of all the bargains struck, says the report, which Mr Annan presented in Cape Town, South Africa, on May 10th.

The report highlights some puzzling details. For instance ENRC, a London-listed Kazakh mining firm, waived its rights to buy out a stake in a mining enterprise owned by Gécamines, Congo's state miner, only to acquire it for $75m from a company owned by Dan Gertler, an Israeli businessman, which had paid $15m for it just months earlier. Mr Gertler is close to Joseph Kabila, Congo's president. ENRC, which is being investigated by the Serious Fraud Office in Britain, was Congo's third-largest copper producer last year. Both ENRC and Mr Gertler deny wrongdoing. African countries often fail to collect reasonable taxes on mining, says Mr Annan's panel. For example, Zambia's copper exports were worth $10 billion in 2011, but its tax receipts from mining were a meagre $240m. The widespread use by mining firms of offshore investment.

Figure 8 The Democratic Republic of Congo

Figure 9 Economic/Political Map of Congo.

This study was a cross-sectional survey conducted between January, and April 2005, in Kinshasa Hinterland. Figure 8 depicts the geographical sketch of the Democractic Republic of Congo Figure 9 shows us the economical/political Map of Congo. Subject recipients of this study were people who were both men and women and who were classified as Bantu. The clustering of cardiovascular risk

factors was defined in all, MS group was done according to IDF (WC, BP, triglycerides, HDL-C, glucose), in the absence and presence of cardio metabolic risk (CDM) group (BMI, WC, BP, fasting glucose, and post-load glucose).

The 107 study sample comprised of participants with cardio metabolic risk and 267 without cardio metabolic risk. In the absence of metabolic syndrome, 2 factors (factor 1=Blood Glucose and factor 2=Obesity) explained 48.1% of the variance. In the presence of metabolic syndrome, three factors were recorded namely; (factor 1=Blood Glucose, factor 2=Blood pressure and factor 3 = Obesity) which explained 73.4% of the variance of the total variance explained.

Out of 977 participants, 17.4 %(n = 170), 11% (n = 107), and 7.7% (n = 75) had type 2 diabetes mellitus (T2DM), MS, and CDM, respectively. Except for BMI, levels of the rest variables were significantly higher in the presence of T2DM than non-diabetics. There was a negative correlation between glucose types and BP in the absence of CDM. Factor analysis among all, BP (factor 1) and triglycerides-HDL (factor 2) explained 55.4% of the total variance, while factor analysis for MS group, triglycerides-HDL-C (factor 1), BP (factor 2), and abdominal obesity-dysglycemia (factor 3) which explained 75.1% of the total variance. In the absence of CDM, glucose (factor 1) and obesity (factor 2) explained 48.1% of the total variance. In the presence of CDM, 3 factors (factor 1 = glucose, factor 2 = BP, and factor 3 = obesity) explained 73.4% of the total variance.

Ethical Considerations

This study was carried out in compliance with the Helsinki Declaration (59th WMA General Assembly, Seoul, South Korea, and October 2008). This research was approved by the Ethics Committee of LOMO Medical Clinic (Ref-00038-03-07) at Kinshasa Limité, Democratic Republic of Congo (DRC). Fully informed and written consent was obtained from each and every adult participant.

Sampling Technique and Sample Size Determination

The survey was specifically and extensively designed using a statistical multistage and stratified random model at each level to recruit a study sample with similar and representative characteristics of Kinshasa Hinterland demographic and socio-economic structure and results were comparable with global data on DM. Each region contributed a number of cluster (EDs) calculated by population number: 185, 112 inhabitants for the upper urban area of Gombe, 161,410 inhabitants of the semi-rural Kisero area, 153,265 inhabitants for the urban Lukemi area and 146,034 inhabitants for the deepest rural Feshi area. The sample size was calculated as: $Z^2 * P * Q/EP$ ($Z^2 xPxQ/EP$) where; Q=1-P, and EP= the expected margin of error.

Variables Included in the Study

The data for this study have been clearly explained under different topics of this thesis and at the beginning of this chapter under the topic of "type of variables". There, some of the variables were stated as; Diastolic Blood Pressure, Systolic Blood Pressure, Waist Circumference, Body Mass Index, Fasting glucose, Post Load Glucose, etc. The study concentrated on these stated variables due to the fact that the original dataset was too large to consider all the variables. This was seen as a way of diluting the analysis and putting less emphasis on important variables. Having selected a few of the variables enabled a more detailed analysis than would have been the case had all the variables been taken into consideration.

Data Preparation: Questionnaire Coding

Responses from both questionnaires were both qualitative and quantitative. Coding was done by assigning numbers to choices or categories for each question. Each number in a record represented a given response. For example, in the first question of the data collection survey, one was assigned to urban areas and two to rural areas. This was done for responses for the whole survey.

This chapter informs the reader about the data used in the study. The first section discusses the data used in the development of Principal Component Analysis. It gives details of the types of variables used followed by discussing the procedures used to include variables in the models. This section discusses the nature of data collected from the survey. Also discussed is the area of study. It further explains the procedure of how the data was collected as well as the sampling process. This chapter concludes with a discussion of how data from the surveys was coded for capturing in

the system. The following section discusses derivation of some variables and as well state and explain the determination of derived variables.

Derived Variables

As stated earlier, two types of data form the data base for this thesis; primary variables and derived variables. Derived variables provide information on values of variables determined from single or combined primary variables. For example; BMI1, BMI2, Fasting Glucose Intolerance, Glucose1, Glucose2, etc. were derived from primary variables. Other derived variables were calculated from a combination of primary variables. As an example, Body Mass Index (BMI) was calculated from the patient's height and Waist Circumference. The combining of tables to form a complete database was necessary in order to compute some derived variables which were obtained from primary variables. Body Mass Index (BMI) was obtained by dividing weight in kg by height in metres squared (kg/m^2). Post Load Glucose in the morning was obtained among participants after 8-12 hours overnight fasting for baseline blood sample and after 2 hour ingesting of 75g glucose to determine plasma glucose. Diabetes Mellitus (DM) presence was coded 1, while DM absence was coded 2 in the univariate analysis.

4.9. Conclusion

This chapter has been devoted to data identification and the procedure used for deriving some data variables such as Body Mass Index (BMI). While the majority of variables in the data set were obtained straight from the respondents, a number of the variables included in the dataset were derived. The researcher has made mention of BMI and others which have been used in this research. Two basic types of data have thus been included in the study. These have been termed by the researcher as; original (primary) variables and derived variables (deduced variables). The analysis was performed by use of the SPSS (Statistical Package for Service Solutions) version 22 & 23.

Chapter 5

5. Data Presentation, Analysis and Interpretation

5.1. Introduction

This chapter presents the analysis and interpretations of the data for this study. This chapter was accomplished by the application of different statistical tools including, tables, figures, and inferential statistics. Due to some limiting factors of analysis and owing to the fact that not all variables would have been included in the analysis, some variables in the data set were not included. Furthermore, there was need for a deeper analysis where emphasis was on a limited number of variables. Thus many variables of interest were examined for their relevance in both exploratory and confirmatory factor analysis to determine their strength for the prediction of the occurrence of metabolic syndrome.

Those variables whose analysis needed more emphasis were; cardio-metabolic factors including: Systolic blood pressure, Diastolic blood pressure, Body mass index, Waist circumference, Glucose fasting and Post load glucose. These variables were identified due to their perceived nature of contribution to cardio metabolic syndrome. The researcher used the method of factor analysis owing to the objectives of this research. It must be emphasized here that the data collection included other variables which could be defined as categorical variables which the researcher classified as independent variables. Thus, the data could easily be classified into two distinct categories namely; dependent variables and independent variables. Independent variables include gender (Male or Female), presence or absence of cardio-metabolic risk, etc. The analysis included the following output:

- Correlation matrix in absence of Cardio-metabolic Risk;
- Correlation Matrix in presence of Cardio-metabolic Risk;
- Rotated Component matrix in absence of Cardio-metabolic Risk;
- Rotated Component Matrix in Presence of Cardio-metabolic Risk;
- Communalities;
- Total Variance Explained;
- Scree Plot ;

- Other analyses include reliability confirmatory factor analysis (CFA). Additionally it explains and elaborates the outcomes of SEM
- Component Matrix.

For all of the stated analyses, Principal Component Factor Analysis was used with the Varimax rotation. The analysis revealed a number of inferences depending on the particular setting of the data. Some similarities existed for analyses based on; gender, presence or absence of cardio metabolic factors, etc. A number of results were observed including the revelation that among all participants, the most important factor to watch for was glucose and neither obesity nor blood pressure was of immediate concern. The analysis follows with explanations according to each setting to be stated in advance

5.2. Exploratory Factor Analysis in the General Population

This section presents analysis, interpretation and discussion on results obtained using exploratory Factor Analysis Several scenarios have been considered with analysis and relevant discussions.

5.2.1. Analysis in the Absence of Cardio-metabolic Risk Descriptive Statistics

The first output from this analysis is recorded in **Error! Reference source not found.** of descriptive statistics for all the variables under investigation. Typically, the mean, standard deviation and number of respondents (N) who participated in the survey are given. The results whose values can be read from the table are self-explanatory as the values are seen in the table. The highest mean (136.87) corresponds Post Load Glucose, followed by Systolic Blood Pressure (117.47) and the lowest observed mean, (1.6466) corresponds to Height in meters. Other variables with their corresponding means have been given in the order of their analysis and inclusion in the data are as follows, Diastolic B Pressure (70.39), Body MI (29.981) Glucose FA (85.25), Waist circumference (79.64). The stated means show that different variables carry different weights in terms of magnitude. The difference in the values of the respective means had no additional meaning except that their overall measurements were equally different.

Table 1 Descriptive statistics in the absence of cardio-metabolic risk in the general population.

	Mean	Std Deviation	Analysis N
Systolic B Pressure	117.47	18.793	374
Diastolic B Pressure	70.39	12.199	374
Height	1.6466	0.13398	374
Body MI	29.981	108.2660	374
Glucose FA	85.25	16.651	374
Waist circumference	79.64	14.125	374

Post Load Glucose	136.87	52.973	374

Correlation Matrix

Error! Reference source not found. shows a matrix of linear correlation between variables of interest in the dataset (Gupta, 2000). A correlation matrix is simply a rectangular array of numbers which gives the correlation coefficients between a single variable and every other variable in the investigation. The correlation coefficient between a variable and itself is always 1, hence the principal diagonal of the correlation matrix contains 1s. The correlation coefficients above and below the principal diagonal are the same. It was observed by the researcher that since so many of the correlations were very small, this would have been a matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0. However, this assumption was rejected by Bartlet's test whose p-values was 0.000…, which meant that the off-diagonal elements of the correlation matrix were not all zero. Further, it was an additional observation that magnitude of the correlations ruled out the possibility of multicollinearity (high correlations between variables). It was observed from table 2 that the following pairs of variables had significant correlations; SBP vs DBP, HT vs BMI and GF vs FLG. It was assumed that these significant correlations (with p less than 0.05) demonstrate the combined effect of the pairs of variables in their contribution to the manifestation of Cardio-metabolic syndrome making their presence known either directly or indirectly. Correlations of other pairs were not significant owing to the high p-values (>0.05).

Table 2 Bivariate correlation matrix in the general population.

	SBP	**DBP**	**HT**	**BMI**	**GF**	**WC**	**PLG**
SBP	1.000	.475	-.007	-.047	-.042	-.006	-.046
DBP	.475	1.000	-.047	.064	-.012	-.011	.008
HT	-.007	-.047	1.000	-.609	.040	-.087	-.002
BMI	-.047	.064	-.609	1.000	.081	.118	.101
GF	-.042	-.012	.040	.081	1.000	.061	.506
WC	-.006	-.011	-.087	.118	.061	1.000	.024
PLG	.046	.008	-.002	.101	.508	.024	1.000

Communalities

Table 3 below is the table of communalities showing how much of the variance in the variables had been accounted for by the extracted factors. For example close to 80% of the variance in Body Mass Index was accounted for while slightly over 74% of the variance in Systolic Blood Pressure has been accounted for by the extracted factors. The method of extraction was Principal Component Analysis. As further interpretation of the results obtained under this topic, it was

observed that close to 80% of the proportion of variation in Body Mass Index was explained by the three extracted factors. It is also summarized from table 3 that the model did better for some variables than it did for others. The model explained better for Body Mass Index, Height, Glucose Fasting, …, Diastolic Blood Pressure but it did not do well for Waist Circumference. The best explanation was given to Body Mass Index.

Table 3 Communalities in the general population.

Variable	Initial	Extraction
Systolic BP	1.000	.743
Diastolic BP	1.000	.735
Height	1.000	.787
Body MI	1.000	.794
Glucose FA	1.000	.757
Waist Circumference	1.000	.087
Post Load Glucose	1.000	.746

Total variance

The next item (**Error! Reference source not found.**) shows all the factors extractable from the analysis along with their eigenvalues, the percent of variance attributable to each factor, and the cumulative variance of the factor and the previous factors. This is shown in the table below. We note here that the first factor accounts for 23.601% of the variance, the second 21.656% and the third 21.150%. All the remaining factors are not significant, and accordingly, they don't count. It was noted here that three factors were extracted by this analysis.

The Scree Plot

The scree plot (i.e. see **Error! Reference source not found.**) is a graph of the eigenvalues against all the factors. The graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 3 and 4. Note also that factor 4 has an eigenvalue of less than 1, so only three factors have been retained. An eigenvalue is the standardized variance associate with a particular factor. The sum of the Eigen-values cannot exceed the number of items in the analysis, since each item contributes one to the sum of variances.

Table 4 Total Variance Explained in the general population.

component	Eigen-value	Eigenvalue % of variance	cumulative	Total Value	% of total	Cumulative %	Rotated Sums of squared loadings	% of Variance	Cum %
1	1.692	24.171	24.171	1.692	24.171	24.171	1.652	23.601	23.601
2	1.496	21.375	45.547	1.496	21.375	45.547	1.516	21.656	45.257
3	1.460	20.861	66.407	1.460	20.861	66.407	1.480	21.150	66.407
4	.966	13.796	80.203						
5	.551	7.868	88.071						
6	.464	6.633	94.705						
7	.371	5.295	100.00						

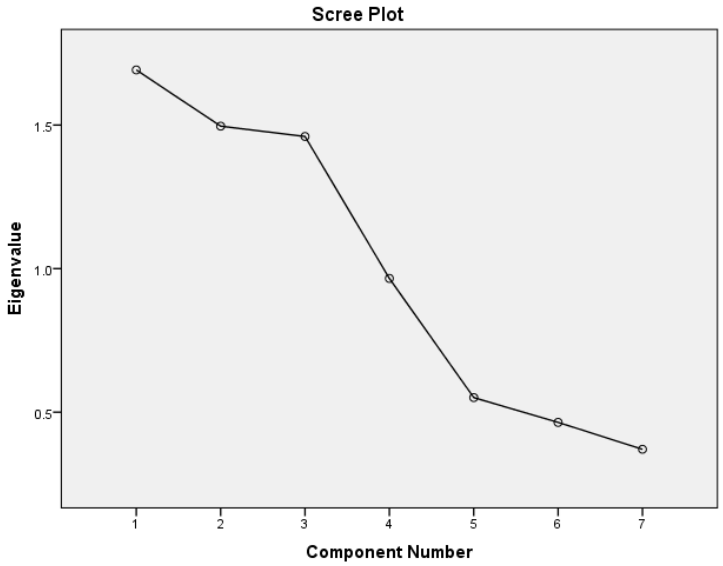

Figure 10 Scree plot for analysis in the general population.

Component (Factor Matrix)

The component (Factor Matrix) contains the loadings of the seven variables on the three factors extracted, (see **Error! Reference source not found.**) for a demonstration of two variable. The higher the absolute value of the loading, the more the factor contributes to the variable. Sometimes, the analysist instructs the computer to leave out some tabular values. Gaps on such a table represent loadings that are less than 0.5, making reading the table an easier task. Thus in some cases, all loadings less than 0.5 could be suppressed.

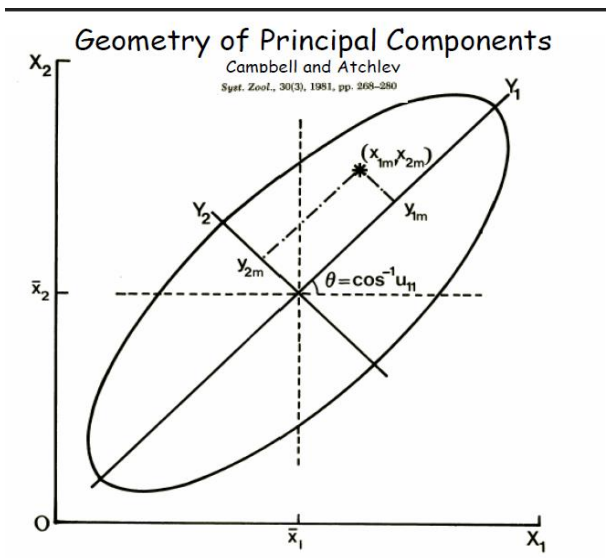

Figure 11 Idealized representation of a scatter diagram for two variables.

Rotated Component (Factor) Matrix

The idea of rotation is to reduce the number factors on which the variables under investigation have high loadings. Rotation does not actually change anything but makes the interpretation of the analysis easier. Looking at the table below (**Error! Reference source not found.**), we can see that Systolic Blood Pressure and Diastolic Blood Pressure are substantially loaded on Factor (Component) 3 while Height and Body Mass Index are substantially loaded on Factor 1. Glucose Fasting and Post Load Glucose are substantially loaded on Factor 2. These factors can be used as variables for further analysis. One big advantage of this type of analysis is their utilization in the completion of CFA.

Table 5 Rotated Component Matrix in the general population.

	Component		
Variable	**1**	**2**	**3**
Systolic BP			.859
Diastolic BP			.856
Height	-.878		
Body MI	.889		
Glucose FA		.868	
Waist Circumference			
Post Load Glucose		.859	

Component Plot in Rotated Space

The plot below (see **Error! Reference source not found.**) shows the items (variables) in the rotated factor space. This was an SPSS output of the research data analysis. According to this figure, three factors were extracted. The logic behind the factor formation is the high correlation of the variables with the respective factors. Considering figure 12, it is noted that the clustering of Height and Body Mass Index formed the first factor "Obesity" or "Overweight" (F1). The second factor "Blood Glucose Metabolism Disordering" (F2) was formed by the clustering of Glucose fasting and Post Load Glucose. The third factor, Factor 3 "Blood Pressure" was formed by the clustering of Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP). The loading observed in table 5 were the loadings which were greater than 0.5 in absolute value.

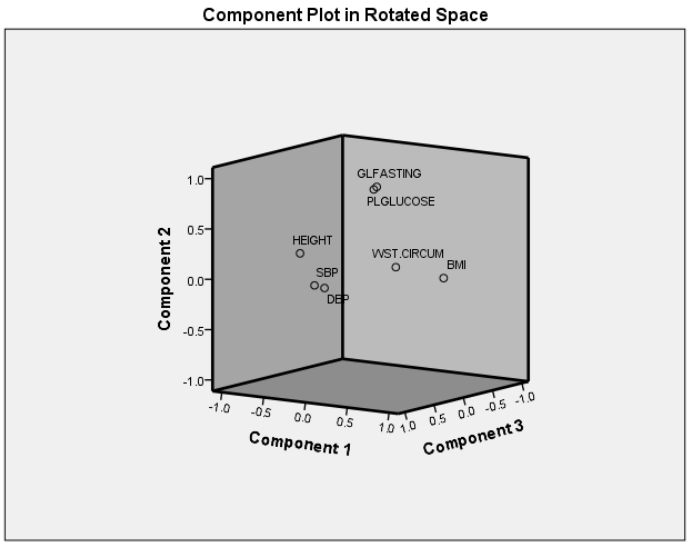

Figure 12 Component Plot in Rotated Space in the general population.

5.2.2. Component Score Coefficient

There are two types of score coefficient statistics. Which are as follows;

- Component Score Coefficient Matrix - This is the factor weight matrix which is used to compute the factor scores, see **Error! Reference source not found.** for an example.

- Component Score Covariance Matrix - because of use of orthogonal rotation making this to be a diagonal matrix, meaning that the same number should appear in all three places along the diagonal. In actual sense, the components are uncorrelated; however, because factor scores are estimated, there is a possibility of the existence of insignificant correlations among the factor scores.

Table 6 Component Score Coefficient Matrix in the general population.

	Component		
Variable	**1**	**2**	**3**
Systolic BP	-.055	.010	.582
Diastolic BP	.017	-.004	.578
Height	-.536	.102	-.021
Body MI	.538	.010	.000
Glucose FA	.000	.572	-.029
Waist Circumference	.165	.054	-.030
Post Load Glucose	.018	.566	.035

5.2.3. Analysis in the Absence of Cardio-metabolic Risk

Table 7 Characteristics of Descriptive Statistics in Absence of Cardio-metabolic Risk.

	Mean	Std. Deviation	Analysis N
Systolic BP	113.41	12.407	267
Diastolic BP	66.56	7.566	267
Height	1.6431	.11301	267
Body MI	24.617	8.8418	267
Glucose FA	81.99	13.973	267
Waist Circumference	79.84	13.552	267
Post Load Glucose	122.06	15.608	267

The correlation matrix in absence of cardio-metabolic risk is presented in **Error! Reference source not found.**. Post-load plasma glucose was significantly and positively correlated to BMI and WC, but significantly and negatively correlated with both SBP and DBP. SBP was significantly and positively correlated with BMI but significantly but negatively correlated with FPG. DBP was significantly and negatively correlated with FPG. Principal Component Factor Analysis revealed two uncorrelated factors that cumulatively explained 48.1% of the observed variance in the absence of Cardio-metabolic risk. The number of those two factors was determined by the scree plot according to Eigen-value (**Error! Reference source not found.**). These two factors could be identified as Blood Glucose Metabolism Disordering (Factor 1; % of variance=26.1) and obesity (Factor 2; % of variance=22).

Table 8 Correlation Matrix In absence of Cardio-metabolic Syndrome.

	SBP	DBP	WT	HT	BMI	GF	WC	PLG
Systolic BP	1	.100	.132	-.019	.113	-.092	.063	-.112
Diastolic BP	.100	1	-.089	-0.15	-.019	-.040	-.021	-.088
Weight	.132	-.089	1	.056	.677	-.110	.087	.005
Height	-.019	-.015	.056	1	-.542	.038	-.136	.030
Body MI	.113	-.019	.677	-.542	1	-.099	.200	.011
Glucose FA	-.092	-.040	-.110	.038	-.099	1	.035	.436
Waist Circ	.063	.021	.087	-.136	.200	.035	1	.103
Post LG	-.112	-.088	.005	.030	.011	.436	.103	1

Table 9 Rotated Component matrix in absence of Cardio-metabolic Risk.

	Factor 1	Factor 2
BMI	-0.131	0.738
SBP	-0.502	0.340
DBP	-0.455	0.114
WC	0.077	0.696
FG	0.765	0.051
Post-Load PG	0.707	0.400

Interpretation

Referring to **Error! Reference source not found.**, we can observe that in the absence of metabolic syndrome, two factors were extracted by the verimax procedure using the Principal component approach. The two factors could be identified by the researcher as: Factor (1); constituted by; fasting glucose with a loading of 0.765, post load glucose with a loading of 0.707 and systolic blood pressure with an absolute value loading of 0.502. Factor (2) was constituted by; body mass index with a loading of 0.738 and waist circumference with a loading of 0.696. From medical records, the two factors were identified as: Glucose and Abdominal Obesity. The interesting research finding here was in the comparison of the number of factors extracted under the presence and absence of metabolic syndrome. Whereas, in the presence there were three factors, only two were established in the absence of MS. The loadings in the absence of MS (for factor (2)) showed that more attention had to be given to abdominal obesity due to heavier loadings as compared to factor (1) loadings.

Construction of the Scree Plot

The steps discussed below are the necessary steps followed to create, interpret and utilize the scree plot.

Determining the Scree Plot

The steps to determine coefficients of a scree plot can has been discussed in chapter 3. The example in the mentioned subsection is a good example", we find the coefficients e_{ij} for a principal component. The solution involves the eigenvalues and eigenvectors of the variance-covariance matrix Σ. Details on determining coefficients can be found in the following source.

Using the Spectral Decomposition Theorem

The variance-covariance matrix can be written as the sum over the *p* eigenvalues, multiplied by the product of the corresponding eigenvector *times* its transpose as We know from theory that the total variation of $\boldsymbol{X}$ to be the trace of the variance-covariance matrix, or alternatively, the sum of the variances of the individual variables. This is also equal to the sum of the eigenvalues. This leads to an interpretation of the components in terms of the amount of the full variation explained by each component. The proportion of variation explained by the i^{th} principal component will then be defined as the eigenvalue for that component divided by the sum of the eigenvalues. In other words, the i^{th} principal component explains the obtained proportion of the total variation.

A related quantity is the proportion of variation explained by the first k principal component. This would be the sum of the first k eigenvalues divided by its total variation. Note that if, the proportion of variation explained by the first k principal components is large, then not much information is lost by considering only the first k principal components.

The Possibility of Reducing the Number of Dimensions

When there exist correlations (multicollinearity) between the x-variables, the data may more or less fall on a line or plane in a lower number of dimensions. For instance, imagine a plot of two x-variables that have a nearly perfect correlation. The data points will fall close to a straight line. That line could be used as a new (one-dimensional) axis to represent the variation among data points. As another example, suppose that we have verbal, math, and total SAT scores for a sample of students. We have three variables, but really (at most) two dimensions to the data because total= verbal +math, meaning the third variable is completely determined by the first two. The reason for saying "at most" two dimensions is that if there is a strong correlation between verbal and math, then it may be possible that there is only one true dimension to the data.

Remark: All of this is defined in terms of the population variance-covariance matrix Σ which is unknown. However, we may estimate Σ by the sample variance-covariance matrix given in the standard formula.

5.2.4. Practical Procedure

Compute the eigenvalues. Then we define our estimated principal components using the eigenvectors as our coefficients. Generally, we only retain the first k principal components. Here we must balance two conflicting desires:

1. To obtain the simplest possible interpretation, we want k to be as small as possible. If we can explain most of the variation just by two principal components then this would give us a much simpler description of the data. The smaller the value of k, the smaller the amount of variation is explained by the first k components.

2. To avoid loss of information, we want the proportion of variation explained by the first k principal components to be large. Ideally as close to one as possible. The log transformation was used to normalize the data.

What we really need to draw our attention to here is the eigenvalues of the variance-covariance matrix. In the SAS output the eigenvalues are in ranked order from largest to smallest. These values have been copied into **Error! Reference source not found.** below for discussion.

Step 1

We examine the eigenvalues to determine how many principal components should be considered;

Table 10 Eigenvalues, and the proportion of variation explained by the principal components.

Component	Eigenvalue	Proportion	Cumulative proportion
1	0.3775	0.7227	0.7227
2	0.0511	0.0977	0.8204
3	0.0279	0.0535	0.8739
4	0.0230	0.0440	0.9178
5	0.0168	0.0321	0.9500
6	0.0120	0.0229	0.9728
7	0.0085	0.0162	0.9890
8	0.0039	0.0075	0.9966
9	0.0018	0.0034	1.000
Total	0.5225		

Referring to Table 10, if you take all of these eigenvalues and sum them and you get the total variance of 0.5223. The proportion of variation explained by each eigenvalue is given in the third column. For example, 0.3775 divided by the 0.5223 equals 0.7227, or, about 72% of the variation is explained by this first eigenvalue. The cumulative percentage explained is obtained by adding the successive proportions of variation explained to obtain the running total. For instance, 0.7227 plus 0.0977 equals 0.8204, and so forth. Therefore, about 82% of the variation is explained by the first two eigenvalues together.

Next we need to look at successive differences between the eigenvalues. Subtracting the second eigenvalue 0.051 from the first eigenvalue, 0.377 we get a difference of 0.326. The difference between the second and third eigenvalues is 0.0232; the next difference is 0.0049. Subsequent differences are even smaller. A sharp drop from one eigenvalue to the next may serve as another indicator of how many eigenvalues to consider. The first three principal components explain 87% of the variation. This is an acceptably large percentage.

Alternative Method to determine the number of principal components is to look at a Scree Plot. With the eigenvalues ordered from largest to the smallest, a scree plot is the plot of $\hat{\lambda}_i$ versus *i.* The number of component is determined at the point, beyond which the remaining eigenvalues are

all relatively small and of comparable size. The following plot (**Error! Reference source not found.**) is made in Minitab.

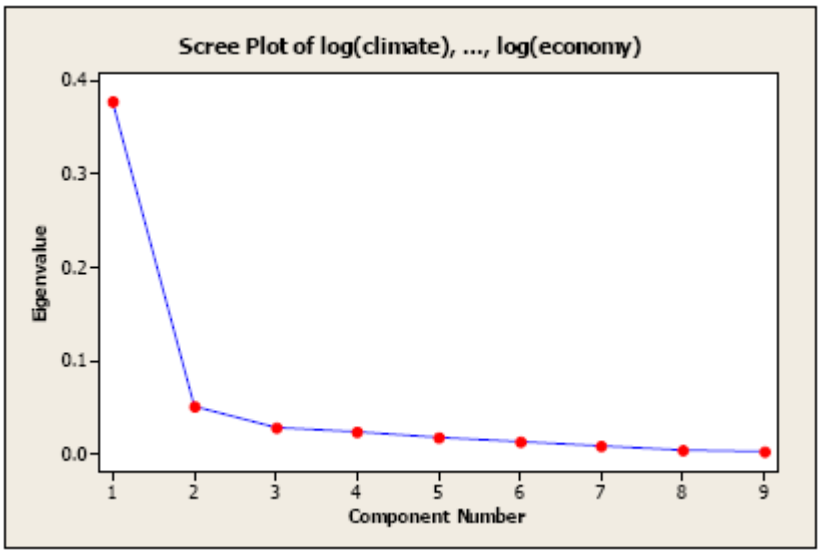

Figure 13 The scree plot for the variables without standardization (covariance matrix).

As you see, we could have stopped at the second principal component, but we continued till the third component. Relatively speaking, contribution of the third component is small compared to the second component and so forth.

Step 2

Next, we will compute the principal component scores. For example, the first principal component can be computed using the elements of the first eigenvector. In order to complete this formula and compute the principal component for the individual communality of interest, plug in that communality's values for each of these variables. A fairly standard procedure is, rather than using the raw data here, to use the difference between the variables and their sample means. This is known as translation of the random variables. Translation does not affect the interpretations because the variances of the original variables are the same as those of the translated variables.

Magnitudes of the coefficients give the contributions of each variable to that component. However, the magnitude of the coefficients also depend on the variances of the corresponding variables.

Interpretation of the Principal Components

Step 3

To interpret each component, we must compute the correlations between the original data for each variable and each principal component. These correlations are obtained using the correlation procedure explained above. In the variable statement we will include the first three principal components, "prin1, prin2, and prin3", in addition to all nine of the original variables. We will use these correlations between the principal components and the original variables to interpret these principal components.

Because of standardization, all principal components will have mean 0. The standard deviation is also given for each of the components and these will be the square root of the eigenvalue. More important for our current purposes are the correlations between the principal components and the original variables. These have been copied into the following table. You will also note that if you look at the principal components themselves that there is zero correlation between the components.

Table 11 Eigenvalues, and the proportion of variation explained by the principal components.

	Principal component		
Variable	**1**	**2**	**3**
Climate	0.190	0.017	0.207
Housing	0.544	0.020	0.204
Health	0.782	-0.605	0.144
Crime	0.365	0.294	0.585
Transportation	0.5858	0.085	0.234
Education	0.394	-0.273	0.027
Arts	0.985	0.126	-0.111
Recreation	0.520	0.402	0.519
Economy	0.142	0.150	0.239

The scree plot

The scree plot (see Figure 14) presents a graph of the eigenvalues against all the factors. The graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 2 and 3. Note also that factor 3 has an eigenvalue of less than 1, so only two factors have been retained. An eigenvalue is the standardized variance associated with a particular factor. The sum of the Eigen-values cannot exceed the number of items in the analysis, since each item contributes one to the sum of variances.

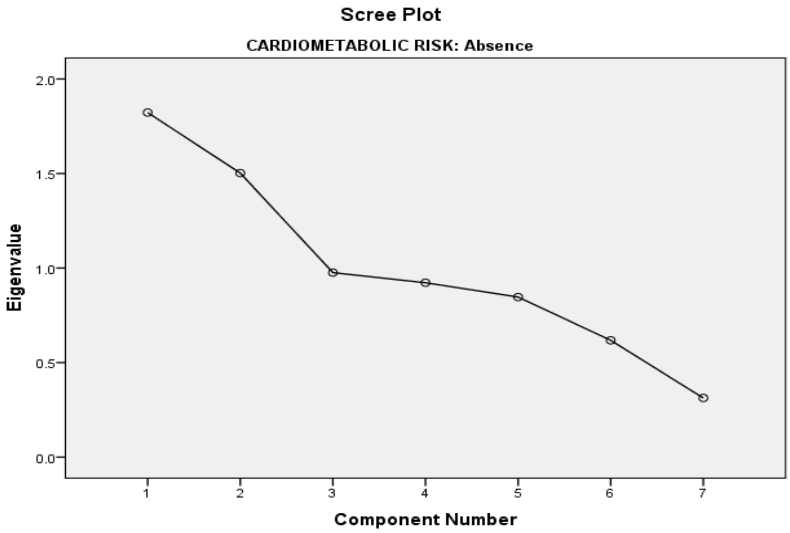

Figure 14 Eigen values among participants in absence of Cardio-metabolic Risk.

Rotation

The next figure shows the factor loadings that result from Varimax rotation. These two rotated factors are just as good as the initial factors in explaining and reproducing the observed correlation matrix. In the rotated factors, Glucose Fasting, Post Load Glucose all have high positive loadings on the first factor, Systolic Blood Pressure, Diastolic Blood Pressure, with high positive loadings on the second factor and Body Mass Index and Waist Circumference having high positive loadings on the third factor.

The next table gives information about the extent to which the factors have been rotated. In this case, the factors have been rotated through 45 degrees. (The angle can be calculated by treating the correlation coefficient as a cosine. The cosine of 45 degrees is .707, see **Error! Reference source not found.**.

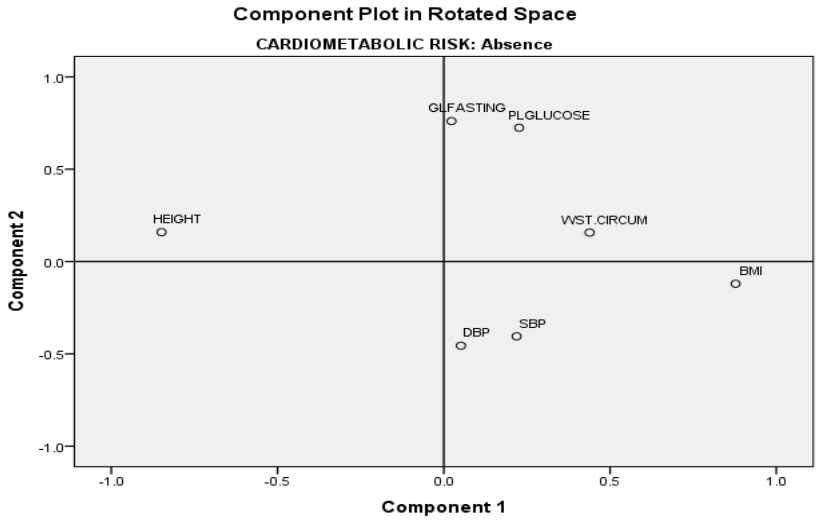

Figure 15 Two-component Plot in Rotated Space among participants in the absence of Cardio -Metabolic Risk.

5.2.5. Factor Analysis in the Presence of Cardio-metabolic Syndrome

Table 12 Descriptive analysis in presence of Cardio-metabolic Risk.

	Mean	Std. Deviation	Analysis N
Systolic Blood Pressure	127.48	27.389	190
Diastolic Blood Pressure	78.02	15.904	190
Weight	64.86	19.356	190
Height	1.6617	.14661	190
Body Mass Index	34.416	151.6317	190
Glucose FA	165.06	144.958	190
Waist Circumference	80.29	15.441	190

Table 12 provides the means of the descriptive statistics for the data in the presence of cardio-metabolic syndrome. From the table, systolic blood pressure has a mean of 127.48 with a standard deviation of 27.389, the mean for diastolic blood pressure is 78.02, the mean for weight is 64.86, the mean for height is 1.6617, the mean for body mass index is 34.416, while glucose has a mean of 165.06 and waist circumference has a mean of 80.29.

This analysis means that the higher the measurement recordings of some variables, the closer the individuals are to developing Cardio-metabolic syndrome. For example, the elevation of: BMI, DBP, SBPP, weight, and waist circumference are pronouncements of expected health problems.

For this setup, Principal Component Factor Analysis (PCFA) extracted three uncorrelated factors that cumulatively explained 73.4% of the observed variance of the presence of Cardio-metabolic risk. The number of these three factors (Components) were determined by the scree plot according to Eigen-values (**Error! Reference source not found.**).

Table 13 below shows a matrix of linear correlation between variables of interest in the dataset. Once again, it was observed that since so many of the correlations were very small, this would have been a matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0. However, this assumption was rejected by Bartlet's test whose p-values was 0.00, which meant that the off-diagonal elements of the correlation matrix were not all zero. Furthermore, it was an additional observation that the magnitude of the correlations ruled out the possibility of multicollinearity (high correlations between variables). It was further observed from table 13 that the following pairs of variables had positive but significant correlations viz Systolic Blood Pressure vs Diastolic Blood Pressure with a positive and significant correlation coefficient and Glucose Fasting vs Post Load Glucose. It was understood that these significant correlations (with p less than 0.05) demonstrated both the combined and individual effects of the pairs of variables in their contribution to the manifestation of Cardio-metabolic syndrome factors and making their presence known either directly or indirectly and as well they demonstrate a direct and significant link with the ethical organizational link with cardio-metabolic factors and the condition itself. Correlations of other pairs were not significant owing to the high p-values (>0.05) obtained.

Table 13 Correlation matrix in presence of Cardio-metabolic Risk.

	Systolic BP	**Diastolic BP**	**Weight**	**Height**	**Body MI**	**Glucose FA**	**Waist Circ.**	**Post LG**
Systolic Blood Pressure	1	.617	-.005	-.049	-.049	-.063	.011	-.140
-.Diastolic Blood Pressure	.617	1	.022	-.065	.022	-.092	-.048	-.264
Height	-.049	-.065	.009	1	-.036	.030	.055	-.011
Body Mass Index	-.049	.022	.045	-.036	1	-.014	.133	.040
Glucose FA	-.063	-.092	.093	.030	-.014	1	.080	.606
Waist Circumference	.011	-.048	.060	.055	.133	.080	1	.060
Post Load Glucose	-.140	-.264	.028	-.011	.040	.606	.060	1

The scree plot

The scree plot (Figure 16) is a graph of the eigenvalues against all the factors. The graph is useful for determining the number of factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 3 and 4. Note also that factor 4 has an eigenvalue of less than 1, so only three factors have been retained. An eigenvalue is the standardized variance associate with a particular factor. The sum of the Eigen-values cannot exceed the number of items in the analysis, since each item contributes one to the sum of variances.

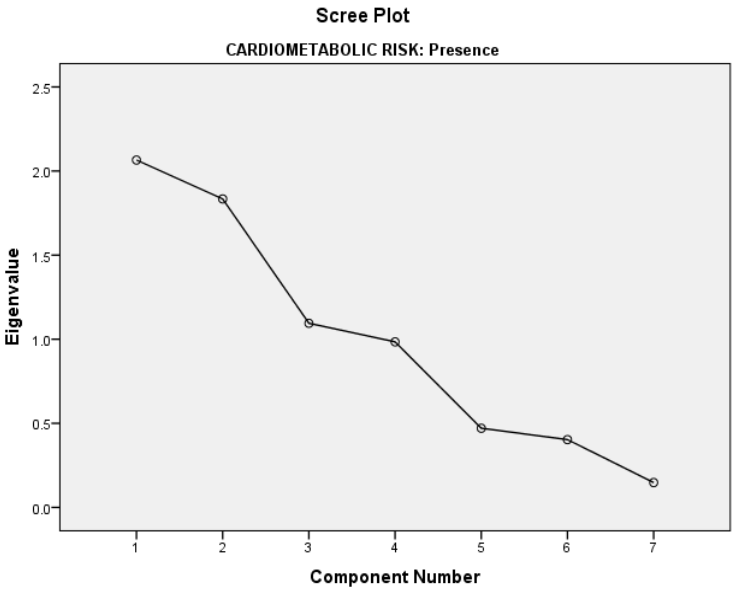

Figure 16 Eigen values among participants with cardio-metabolic risk.

Table 14 Rotated Component Matrix in Presence of Cardio-metabolic Risk.

	Factor 1	Factor 2	Factor 3
BMI	.077	.025	.760
SBP	-.064	.881	-.137
DBP	-.0224	.833	.116
WC	.025	.001	.769
FPG	.906	-.106	.096
Post Load PG	.894	-.179	.031

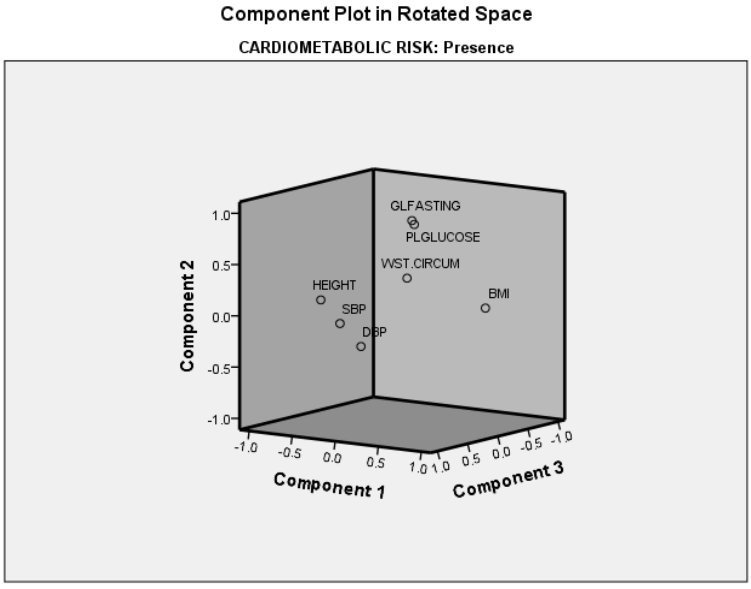

Figure 17 The three-component plot in Rotated Space among participants.

Figure 17 shows the items (variables) in the rotated factor space. This was an SPSS output of the research data analysis. According to this figure, three factors were extracted. The logic behind the factor formation is the high correlation of the variables with the respective factors. It is noted that the clustering of FPG Post Load PG formed the first factor "Obesity" or "Blood Glucose Metabolism Disordering" (F1). The second factor "Blood Pressure" (F2) was formed by the clustering of SBP and DBP. The third factor, Factor 3 "Obesity" was formed by the clustering of BMI and waist circumference (WC). The loading observed in table 14 were the loadings which were greater than 0.5 in absolute value.

5.2.6. Male Participants in the Presence of Cardio-metabolic Risk

Descriptive Statistics

The descriptive output suggests that among all the variables included in the study, it was clear that the most important among them that influence the presence of cardio-metabolic risk are; Post load glucose with a mean of 173.81 and a standard deviation of 85.632, systolic blood pressure with a mean of 127.62 and a standard deviation of 26.665, glucose fasting with a mean of 93.41 and a standard deviation of 19.787 and diastolic blood pressure with a mean of 79.95 and a standard deviation of 15.834. The other variables have comparatively smaller means, see **Error! Reference source not found.**.

Table 15 Descriptive Statistics for males in the presence of Metabolic Syndrome.

Variable	Mean	Std. Deviation	Analysis N
Systolic Blood Pressure	127.62	26.665	107
Diastolic Blood Pressure	79.95	15.834	107
Height	1.6552	0.17610	107
Body Mass Index	43.367	201.9827	107
Glucose FA	93.41	19.787	107
Waist Circumference	79.16	15.520	107
Post Load Glucose	173.81	85.632	107

Table 16 Correlation matrix for males in the presence of cardio metabolic risk.

Variable	Systolic BP	Diastolic BP	Height	weight	Body MI	Glucose FA	Waist Circ.	Post LG
Systolic Blood Pressure	1	.559	.001	.054	-.073	-.011	.007	-.134
Diastolic Blood Pressure	.559	1	.045	-.046	.034	-.085	-.050	-.304
Height	.001	.045	1	.086	.034	.014	.060	-.093
Weight	.054	-.046	.086	1	-.747	.057	.056	-.066
Body Mass Index	-.073	.034	.034	-.747	1	-.026	.133	.052
Glucose FA	-.011	-.085	.014	.057	-.026	1	.082	.533
Waist Circumference	.007	-.050	.060	.056	.133	.082	1	.060
Post Load Glucose	-.134	-.304	-.093	-.066	.052	.533	.060	1

Correlation Matrix

The above output from the analysis is the correlation coefficient r in the matrix for males in the presence of cardio-metabolic syndrome (see **Error! Reference source not found.**). The results shows a matrix of linear correlations between observed variables in the dataset. Once again, it was observed that since so many of the correlations were very small, this would have been a matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0. However, this assumption was rejected by Bartlet's test whose p-values was 0.00, which meant that the off-diagonal elements of the correlation matrix were not all zero and in fact none was 0.00. It was an additional observation that magnitude of the correlations ruled out the possibility of multicollinearity (high correlations between variables). Also noted from table 16 was pairs of variables which had significant correlations; Systolic Blood Pressure vs Diastolic Blood Pressure having a significant correlation of 0.559 (P<0.01), and Glucose Fasting vs Post Load Glucose with a correlation coefficient of 0.533 (P-value < 0.01). It was assumed that these significant correlations (with p less than 0.05) had a positive contribution to the manifestation of Cardio-metabolic syndrome, making their presence felt either directly or indirectly. Correlations of other pairs were not significant owing to their negligible correlation coefficients with high p-values (>0.05).

Communalities

The communalities for the i^{th} variable are computed by taking the sum of the squared loadings for that variable. This is expressed below:

$$\hat{h}_i = \sum_{j-1}^{m} \hat{l}_{ij}^2 \tag{51}$$

To understand the computation of communalities, consider the following table of factor loadings according to the listed variables:

Table 17 Factor loadings according to the listed variables.

	Factor		
Variable	**1**	**2**	**3**
Climate	0.286	0.076	0.841
Housing	0.698	0.153	0.084
Health	0.744	-0.410	-0.020
Crime	0.471	0.522	0.135
Transportation	0.681	-0.156	-0.148
Education	0.498	-0.498	-0.253
Arts	0.861	-0.115	.-0.011
Recreation	0.642	0.322	0.044
Economics	0.298	0.595	-0.533

For example, to compute the communality for Climate, the first variable, we square the factor loadings for climate (given in bold-face in the table above) then add the results:

$$\widehat{h}_1 = 0.28682^2 + 0.07560^2 + 0.84085^2 = 0.7950$$

The communalities of the 9 variables can be obtained as shown in the example for climate.

Table 18 Communalities of the 9 variables.

	Final Communality Estimates: Total=5.616885							
Climate: 0.7950		Housing: 0.5178		Health: 0.7223		Crime: 0.5124		Trans: 0.5098
	Educate: 0.5607		Arts: 0.7538		Recreate: 0.5173		Econ: 0.7277	

From **Error! Reference source not found.**, we can see that 5.616885 (located just above the individual communalities), is the "Total Communality". In summary, the communalities are stated in the table (table 19) below

Table 19 Communalities for 9 listed variables.

Variable	**Communality**
Climate	0.795
Housing	0.518
Health	0.722
Crime	0.512
Transportation	0.510
Education	0.561
Arts	0.754
Recreation	0.517
Economics	0.728
Total	5.617

These may be taken to be multiple R^2 values for regression models predicting the variables of interest from the 3 factors. The communality for a given variable can be interpreted as the proportion of variation in that variable explained by the three factors. In other words, considering results in *table 19* if we perform multiple regression of climate against the three common factors, we obtain an R^2= 0.795, Table 19 indicates that about 80% of the variation in climate is explained by the factor model.

The results suggest that the factor analysis does the best job of explaining variation in climate, the arts, economics, and health. One assessment of how well this model is doing can be obtained from the communalities. What you want to see is values that are close to one. This would indicate that the model explains most of the variation for those variables. In this case, the model does better for some variables than it does for others. The model explains Climate the best, and is not bad for other variables such as Economics, Health and the Arts. However, for other variables such as Crime, Recreation, Transportation and Housing the model does not do a good job, explaining only about half of the variation.

If you take all of the communality values and add them up you can get a total communality value:

$$\sum_{i-1}^{p} \hat{h}_i = \sum_{i-1}^{m} \hat{\lambda}_i \tag{52}$$

Here, the total communality is 5.617. The proportion of the total variation explained by the three factors. This gives us the percentage of variation explained in our model. This might be looked at as an overall assessment of the performance of the model. However, this percentage is the same as the proportion of variation explained by the first three eigenvalues, obtained earlier. The individual communalities tell how well the model is working for the individual variables, and the total communality gives an overall assessment of performance. These are two different assessments that you can use.

Since the data are standardized in this case, the variance for standardized data is going to be equal to one. Then the specific variances can be computed by subtracting the communality from the variance as expressed below:

$$\hat{\psi}_i = 1 - \hat{h}_i \tag{53}$$

Recall, that in (Table 17 and Table 19) the data were standardized before analysis, so the variances of the standardized variables are all equal to one. For example, the specific variance for Climate is computed as follows:

$$\hat{\psi}_1 = 1 - 0.795 = 0.205$$

$$\frac{5.617}{9} = 0.624$$

Communalities

The following analysis and interpretation refers to communalities for the variables stated in the analysis. The table of communalities shows how much of the variance in the variables was accounted for by the extracted factors. For instance, from the table of communalities below, it is noted that 75.6% of the variance in systolic blood pressure is accounted for while 70.6% of the variance in Post Load Glucose is accounted for. On the other hand the researcher noted that 76.1% of the variance in Glucose Fasting is accounted for. Other variance accountings are as stated in the following table of communalities, (see **Error! Reference source not found.**). Further, it is understood that the extraction of the factors explained 92.1% of the proportion of variance in Body Mass Index, 89.3% of the proportion of variation in Height, 76.1% in Glucose Fasting, 75.9% in Diastolic Blood Pressure, etc. For these variables, the model does comparatively better that other variables such as: Waist Circumference for which only 19.9% of the proportion of variation was explained by the three extracted factors, Post Load Glucose for which 70.6% of its proportion of variation was explained by the extracted factors.

Table 20 table of Communalities for Males in the Presence of Cardio-metabolic Risk.

Variable	Initial	Extraction
Systolic Blood Pressure	1.000	.756
Diastolic Blood Pressure	1.000	.759
Height	1.000	.893
Body Mass Index	1.000	.921
Glucose FA	1.000	.761
Waist Circumference	1.000	.199
Post Load Glucose	1.000	.706

Total Variance Explained

Table 21 below shows all the factors extractable from the analysis along with their Eigen values. Furthermore, the percent of variance attributable to each factor and the cumulative variance of the factor for various previous factors are shown by use of Principal Component Analysis. Notice that the first factor accounts for 26.56% of the variance, the second 22.511% and the third 22.276%. All the remaining factors are not significant. This conclusion is based on the logic that the total variance for each of the factors is less than 1.00.

The Scree Plot in males in the Presence of Cardio-Metabolic Syndrome

Figure 18 below shows the scree plot for males in the presence of cardio-metabolic syndrome. By the definition of the use of the scree plot to decide on the number of factors, it is easily read form the figure that the analysis extracted three factors for this setting. The arm in the figure starts to show a significant bent at the three factor level. The rest of the factors are not significant.

Table 21 Total variance explained for males in the presence of Metabolic Risk.

Initial Eigenvalues				**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative**	**Total**	**% of variance**	**Cumulative %**
1	2.065	29.506	29.506	2.065	29.506	29.506	1.859	26.560	26.560
2	1.834	26.202	55.708	1.834	26.202	55.708	1.576	22.511	49.071
3	1.095	15.639	71.347	1.095	15.639	71.347	1.559	22.276	71.347
4	.984	14.061	85.408						
5	.471	6.723	92.130						
6	.403	5.759	97.890						
7	.148	2.110	100.00						

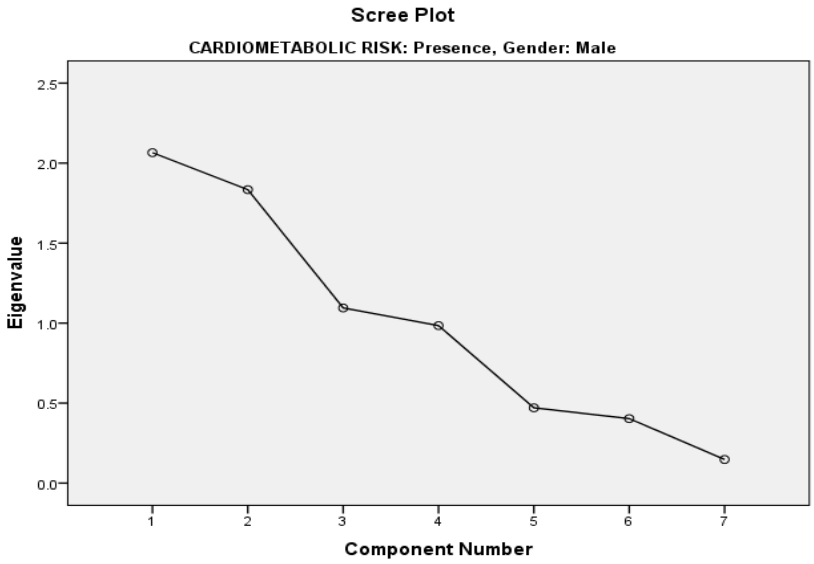

Figure 18 Scree Plot for Males in Presence of Metabolic Syndrome.

The scree plot

The scree plot in Figure 18 above is a graph of the eigenvalues against all the factors. The graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 3 and 4. Note also that factor 4 has an eigenvalue of less than 1, so only three factors have been retained. An eigenvalue is the standardized variance associate with a particular factor. The sum of the Eigen-values cannot exceed the number of items in the analysis, since each item contributes one to the sum of variances.

Component Matrix

The table below shows the loadings of the six variables on the three factors extracted. The higher the absolute value of the loading, the more the factor contributes to the variable. This analysis included all loadings even those that were less than 0.5. It is easily observed that the three factors can be identified as follows 3 components:

- Comprising of the variables; Glucose fasting with a loading of 0.734 and Post load glucose with a loading of 0.731. Medically these two variables can be grouped to form the factor of Sugar (glucose: DM);
- Comprising of the variables; Waist Circumference with a loading of 0.708 and Body Mass Index with a loading of 0.687. These two variables are medically grouped under the overall factor of Obesity;
- Comprising of Diastolic Blood Pressure with a loading of -0.723 and finally and Systolic Blood Pressure with a loading of -0.617. These two variables can be easily classified as Blood Pressure (Hypertension).

It is understood here that the most important variable for this factor analysis that contributes to the influence of metabolic syndrome is identified to be Glucose. This means that as far as the research is concerned, sugar is the most important among all metabolic factors. This is followed quite closely by Obesity. This is contrary to earlier observed beliefs that blood pressure is the most important among metabolic factors when observed in the presence of Cardio metabolic Risk.

Rotated Component Matrix

Error! Reference source not found. gives the rotated component matrix for the component matrix analyzed above. The main advantage of rotation is to make the interpretation of the analysis easier. Looking at the above interpretation, it was not easy to create the three factors as separately as it was required. There is a sort of mix of the variables where you find that some factors were found intertwined less than one component. The following table brings out a clearer component analysis. Of course, it must be understood that the higher the absolute value of the loading, the more the factor contributes to the variable.

As noted above, applying the same logic here, the first component is composed of two well defined variables; Height with a loading of 0.923 and Body Mass Index with a loading of -0.935. These two clearly form the factor which is identified as *obesity*. The second component comprises of Systolic Blood Pressure with a loading of 0.855 and Diastolic Blood Pressure with a loading of 0.886. These two variables constitute to the second factor of *Blood Pressure*. Component three comprises of Glucose Fasting with a loading of 0.657 and Waist Circumference with a loading of 0.781. These two variables form the factor which is easily classified as *Obesity*. This is interpretation for factor analysis for the six variables under this study.

Table 22 Component Matrix for males in the presence of Metabolic Syndrome.

	Component		
Variable	**1**	**2**	**3**
Systolic Blood Pressure	.164	.855	.177
Diastolic Blood Pressure	.017	.886	.038
Height	.923	-.097	.087
Body Mass Index	-.935	.061	.107
Glucose FA	.086	-.176	.657
Waist Circumference	-.086	-.087	.781

Table 23 Rotated component matrix for males in the presence of Cardio-metabolic Risk.

	Component		
Variable	**1**	**2**	**3**
Systolic Blood Pressure	.081	.884	.044
Diastolic Blood Pressure	-.066	.879	-.098
Height	.927	.004	.105
Body Mass Index	-.928	-.012	.093
Glucose FA	.090	-.064	.676
Waist Circumference	-.092	.026	.785

The table below gives a Coefficient Score Coefficient Matrix. This explains more in detail and brings out more clarity about the extraction of different factors from the variables. From the Score Coefficient Matrix, one notices that the first component extracted was formed by the most

substantially loaded variables namely; Height with a loading of -0.512 and Body Mass Index with a loading of 0.510. The second component is composed of Glucose Fasting with a loading of 0.560 and Post Load Glucose with a loading of 0.519. The third component comprises of two variables Systolic Blood Pressure and Diastolic Blood Pressure loaded with 0.605 and 0.542 respectively.

Table 24 Component Score Coefficient Matrix for Males In the Presence of CMR.

	Component		
Variable	**1**	**2**	**3**
Systolic Blood Pressure	-.078	.177	.605
Diastolic Blood Pressure	.055	.018	.542
Height	-.512	.053	.018
Body Mass Index	.510	.035	-.005
Glucose FA	-.044	.560	.051
Waist Circumference	.080	.283	.200
Post Load Glucose	-.044	.519	.008

Males without Cardio Metabolic Risk

Descriptive Statistics

The descriptive output suggests that among all the variables included in the study, it is clear that the most important among them that influence the presence of cardio metabolic risk are; Post load glucose with a mean of 122.11 and a standard deviation of 15.688, systolic blood pressure with a mean of 113.33 and a standard deviation of 12.435, glucose fasting with a mean of 82.00 and a standard deviation of 13.985, waist circumference with a mean of 79.76 and a standard deviation of 13.603. The other variables have comparatively smaller means, (see **Error! Reference source not found.**).

Table 25 Descriptive statistics for males in the absence of CMR.

Variable	**Mean**	**Std. Deviation**	**Analysis N**
Systolic Blood Pressure	113.33	12.435	264
Diastolic Blood Pressure	66.59	7.597	264
Height	1.6425	.11344	264
Body Mass Index	24.588	8.8829	264
Glucose FA	82.00	13.985	264
Waist Circumference	79.76	13.603	264
Post Load Glucose	122.11	15.688	264

The Correlation Matrix

Error! Reference source not found. shows a matrix of correlation coefficients among males in the absence of cardio-metabolic syndrome. It further shows a matrix of linear correlations between some observed variables in the dataset. Once again, it was observed that since so many of the

correlations were very small, this would have been a matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0. However, this assumption was rejected by Bartlet's test whose p-values was 0.00, which meant that the off-diagonal elements of the correlation matrix were not all zero and in fact none was 0.00. It was an additional observation that the magnitude of the correlations ruled out the possibility of multicollinearity (high correlations between variables). Also noted from table 26 was pairs of variables which had significant correlations; Body Mass Index vs Height having a significant correlation of -.677 (P<0.01), and Glucose Fasting vs Post Load Glucose with a correlation coefficient of 0.372 (P-value < 0.05). Body Mass Index vs Weight with a p-value of (P<0.05) It was assumed that these significant correlations (with p less than 0.05) had a impact on the existence of Cardio-metabolic syndrome. Correlations of other pairs were not significant owing to their negligible correlation coefficients with high p-values (>0.05).

Table 26 Correlation matrix for males in the absence of cardio-metabolic risk.

Variable	Systolic BP	Diastolic BP	Height	Body MI	Glucose FA	Waist Circ.	Post Load Glucose
Systolic Blood Pressure	1	.132	-.125	.100	-.156	.061	-.057
Diastolic Blood Pressure	.132	1	-.062	.055	-.122	-.020	-.126
Height	-.125	-.062	1	-.677	.051	-.140	-.005
Body Mass Index	.100	.055	-.677	1	-.086	.200	.074
Glucose FA	-.156	-.122	.051	-.086	1	.033	.372
Waist Circumference	.061	-.020	-.140	.200	.033	1	.105
Post Load Glucose	-.057	-.126	-.005	.074	.372	.105	1

Communalities

The following analysis and interpretation in **Error! Reference source not found.** refers to communalities for the variables in the analysis under this topic. The table of communalities shows how much of the variance in the variables has been accounted for by the extracted factors. For instance, from the table of communalities below, it is noted that 57.50%% of the variance in Post Load Glucose is accounted for while 57.9% of the variance in Glucose Fasting is accounted for. On the other hand the researcher noted that 21.80% of the variance in Waist Circumference is accounted for. Other variance accountings are as stated in the following table of communalities. The accounting of the variables in the table in question below, they define how well the model was doing, implying that the model was doing better for some variables than for others. For example, the proportion of the variation in Body Mass Index was explained by the three factors, while 74.8% of the proportion of the variation in height was explained by the three extracted factors, 57.9% of the variation in Glucose Fasting was explained by the three factors. The significance of the

proportion of variation as explained by the observed factors lies in the fact that the observation of the factors generated sufficient information to explain the observed percentage of the original variable as observed form the original data. The quantification of proportion of variance explained is the variance.

Table 27 Communalities for males in the absence of cardio-metabolic risk.

	Initial	Extraction
Systolic Blood Pressure	1.000	.212
Diastolic Blood Pressure	1.000	.216
Height	1.000	.748
Body Mass Index	1.000	.784
Glucose Fasting	1.000	.579
Waist Circumference	1.000	.218
Post Load Glucose	1.000	.575

Total Variance Explained

Error! Reference source not found. shows all the factors extractable from the analysis along with their Eigen values. Furthermore, the percent of variance attributable to each factor and the cumulative variance of the factor for various previous factors are also shown. Notice that the first factor accounts for 26.072% of the variance and the second factor accounts for 22.072%. All the remaining factors are not significant. Like in the earlier analysis, this conclusion is based on the basis that the total variance for each of the other factors is less than 1.00.

Table 28 Total Variance Explained for males in the absence of cardio-metabolic risk.

Initial Eigenvalues				**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**		
Component	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative**
1	1.825	26.072	26.072	1.825	26.072	26.072	1.878	25.535	25.535
2	1.507	21.535	47.607	1.507	21.535	47.607	1.545	22.072	47.607
3	.978	13.969	61.576						
4	.918	13.113	74.690						
5	.846	12.089	86.779						
6	.615	8.785	95.563						
7	.311	4.437	100.00						

The Scree Plot

The scree plot is a graph of the eigenvalues against all the factors (i.e. see **Error! Reference source not found.**). The graph is useful for determining how many factors to retain. The point of

interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 2 and 3. Note further that factor 3 has an Eigen value of less than 1, so only two factors have been retained.

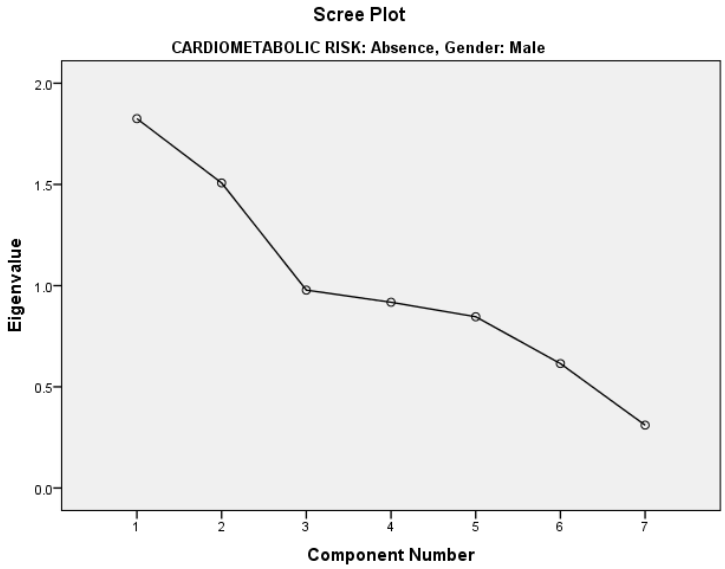

Figure 19 Scree plot for males in the absence of cardio metabolic risk.

Component Matrix

Error! Reference source not found. shows the loadings of the six variables on the two factors extracted. This analysis is quite different from the previous one where there were three variables. The higher the absolute value of the loading, the more the factor contributes to the variable. This analysis included all loadings even those that were less than 0.5. It is easily observed that the two factors can be identified as follows two components:

- Comprising of the variables; Glucose fasting with a loading of 0.766 and Post load glucose with a loading of 0.709. Medically these two variables can be grouped to form the factor of Sugar (glucose);
- Comprising of the variables; Body Mass Index with a loading of 0.727 and Waist Circumference with a loading of 0.698. These two variables are medically grouped under the overall factor of Obesity.

It is understood here that the most important variable for this factor analysis that contributes to the influence of cardio metabolic risk is identified to be Glucose. This means that as far as the research is concerned, sugar is the most important among all cardio metabolic factors. This is followed quite closely by Obesity. This is contrary to earlier observed scientific medical beliefs that blood pressure is more important among cardio metabolic risk factors than other variables in the absence of Cardio metabolic Risk. This research has discovered that contrary to earlier theories and findings, the most important factor is the sugar level in the participant and not necessarily the reading of blood pressure be it systolic or diastolic.

Table 29 Component matrix for males in the absence of cardio-metabolic syndrome.

	Component	
Variable	**1**	**2**
Systolic Blood Pressure	.342	-.308
Diastolic Blood Pressure	.208	-.416
Height	-.853	-.142
Body Mass Index	.865	.188
Glucose Fasting	-.239	.723
Waist Circumference	.360	.297
Post Load Glucose	-.034	.757

Error! Reference source not found. gives the rotated component matrix for the component matrix analyzed above. The main advantage of rotation is to make the interpretation of the analysis easier. The following table brings out a clearer component analysis. Of course, it must be understood that the higher the absolute value of the loading, the more the factor contributes to the variable. As noted, applying the same logic here, the first component; component 1 is composed of two well defined variables; Glucose fasting with a loading of 0.766 and Post Load Glucose with a loading of 0.714. These two clearly form the factor which is identified as Glucose (Sugar).

Component 2 comprises of Body Mass Index with a loading of 0.728 and Waist Circumference with a loading of 0.696. These two variables form the factor which is easily classified as *Obesity.* This is the interpretation for factor analysis for the six variables under this study. This is with specific reference to the case where the research is focused on data collected for participants in the absence of cardio-metabolic risk.

Table 30 Rotated component matrix for males in the absence of cardio-metabolic syndrome.

	Component	
Variables	**1**	**2**
Systolic Blood Pressure	.215	-.407
Diastolic Blood Pressure	.053	-.462
Height	-.580	.160
Body Mass Index	.877	-.121
Glucose Fasting	.024	.761

Waist Circumference	.440	.155
Post Load Glucose	.228	.723

Error! Reference source not found. shows the above analysis but in a diagrammatic form where the six variables in the analysis are grouped according to the pair's common factor. It is noted that similar to the above observation, the variables are grouped as: Component 1 (Post Load Glucose and Glucose Fasting; factor one), component 2 (Systolic Blood Pressure and Diastolic Blood Pressure; factor two) and component 3 (Body Mass Index and Waist Circumference, factor three) (See plot below):

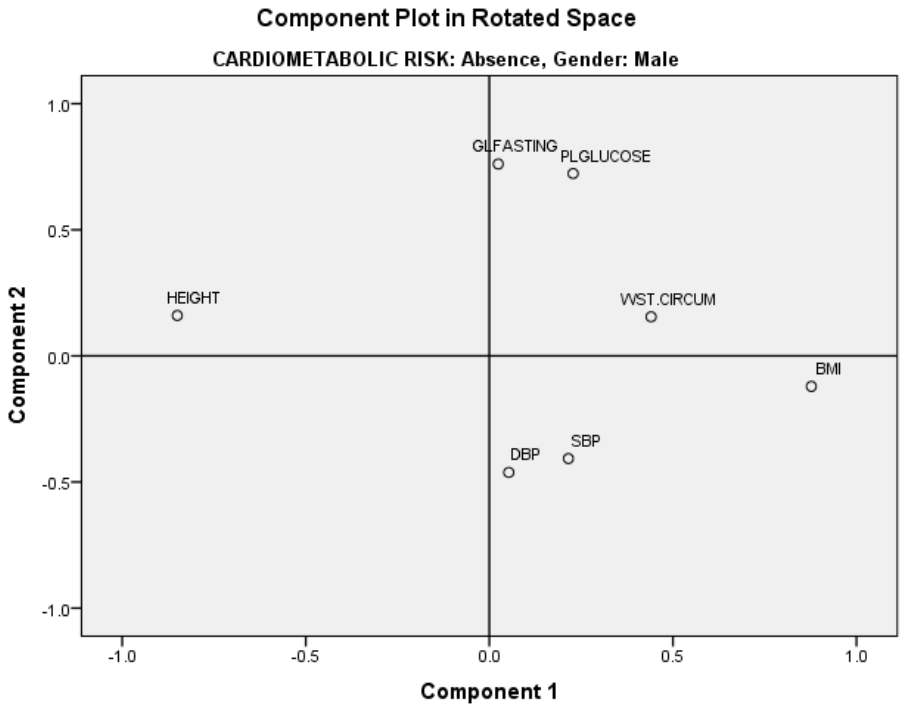

Figure 20 Component plot in rotated space for males I the absence of Cardio-metabolic risk.

5.2.7. Findings in Females without Cardio-metabolic Risk

Descriptive Statistics

For the analysis under this sub topic, the descriptive output suggests that among all the variables included in the study, it is clear that the most important among them that influence the presence of cardio-metabolic risk are; Systolic Blood Pressure with a mean of 120.00 and a standard deviation of 8.660, followed by Post Load Glucose with a mean of 117.67 and a standard deviation of 2.517, Waist Circumference with a mean of 87.00 and a standard deviation of 4.359The fourth in the order of importance is Glucose Fasting with a mean of 80.33 and a standard deviation of 15.631.

The other variables have comparatively smaller means. The means and standard deviation for all the variables are stated in **Error! Reference source not found.**.

Table 31 Descriptive for females in the absence of cardio-metabolic syndrome.

	Mean	Standard Deviation	Analysis N
Systolic Blood Pressure	120.00	8.660	3
Diastolic Blood Pressure	63.33	2.887	3
Height	1.6900	.05568	3
Body Mass Index	27.206	3.3721	3
Glucose Fasting	80.33	15.631	3
Waist Circumference	87.00	4.359	3
Post Load Glucose	117.67	2.517	3

Correlation Matrix

Table 32 shows a matrix of linear correlations between some observed variables in the dataset. Once again, it was observed that since so many of the correlations were very small, this would have been a matrix in which all of the diagonal elements are 1 whereas all off diagonal elements are 0. However, this assumption was rejected by Bartlet's test whose p-values was 0.00, which meant that the off-diagonal elements of the correlation matrix were not all zero and in fact none was 0.00. It was an additional observation that the magnitude of the correlations ruled out the possibility of multicollinearity (high correlations between variables). Also noted from table 32 was that some pairs of variables had significant correlations; Body Mass Index vs Waist Circumference having a negative but significant correlation of -.666 ($P < 0.01$), Body Mass Index vs Height with a negative but significant correlation of -0.356 and a p-value of less than 0.05, Waist Circumference vs Systolic Blood Pressure with a negative but significant correlation of -0.397 have an indirect link with organizational ethical relatedness with the cardio-metabolic disease symptoms that influence individual attack by metabolic syndrome factors. These two signify an indirect link with the factors of cardio-metabolic diseases. Glucose Fasting vs Post Load Glucose with a significant positive correlation and a ($P - value < 0.05$). The significant positive correlation of 0.480 has a direct and combined link with the existence of metabolic factors in the body. The p-value was less than 0.05. Body Mass Index vs Weight with a p-value of ($P < 0.05$) It was assumed that these significant correlations (with p-values less than 0.05) had a direct impact on the existence of Cardio-metabolic syndrome factors. The significant positive correlations demonstrate a high degree of association between the pairs which signify a direct internal organizational derivation of metabolic factors in an individual. In addition, waist circumference and glucose fasting with a significant but positive correlation of 0.859 (p-values less than 0.001, which shows a high level of association between the two variables. This high but significant association shows a direct link with cardio-metabolic factors which are strongly related to one another. The significant positive correlations show that the

individual variables forming the pairs of strongly associated variables have equally significant associations with the underlying cardio-metabolic condition of an individual. The uncorrelated pairs of variables have little or no effect on existence of cardio-metabolic syndrome factors.

Table 32 Correlation matrix for females in the absence of cardio-metabolic risk.

	Systolic BP	Diastolic BP	Height	Body MI	Glucose FA	Waist Circ.	Post Load Glucose
Systolic Blood Pressure	1	.073	.087	.131	-.037	-.397	-.149
Diastolic Blood Pressure	.073	1	.031	-.105	.030	.993	-.059
Height	.087	.031	1	-.356	.027	.700	.060
Body Mass Index	.131	-.105	-.356	1	-.116	-.666	-.046
Glucose Fasting	-.037	.030	.027	-.116	1	.859	.480
Waist Circumference	-.397	.993	.700	-.666	.859	1	-.228
Post Load Glucose	-.149	-.059	.060	-.046	.480	-.228	1

Communalities

The following analysis and interpretation refers to communalities for the variables in the analysis. **Error! Reference source not found.** of communalities shows how much of the variance in the variables has been accounted for by the extracted factors. For instance, from the table of communalities below, it is noted that 100.00% of the variance in systolic blood pressure is accounted for while 100.00% of the variance in Post Load Glucose is accounted for. On the other hand the researcher noted that 100.00% of the variance in Glucose Fasting is accounted for. In a similar rating, other variance accountings are as stated in **Error! Reference source not found.**. Accounting in this case has a variance implication where 100% accounting meant that 100% of the proportion of variation in all the seven recorded variables was explained by the factors. The understanding here by explained variation was the measures of the proportion to which a mathematical model accounts for the variation (dispersion) of a given data set.

Table 33 Communalities for females in the absence of cardio-metabolic risk.

	Initial	Extraction
Systolic Blood Pressure	1.000	1.000
Diastolic Blood Pressure	1.000	1.000
Height	1.000	1.000
Body Mass Index	1.000	1.000
Glucose Fasting	1.000	1.000
Waist Circumference	1.000	1.000
Post Load Glucose	1.000	1.000

Error! Reference source not found. shows all the factors extractable from the analysis along with their Eigen values. Furthermore, the percent of variance attributable to each factor and the

cumulative variance of the factor for various previous factors are also shown. Notice that the first factor accounts for 58.182% of the variance while the second factor accounts for 41.818%. All the remaining factors are not significant. This conclusion is based on the logic that the total variance for each of the factors is less than 1.00. Only cases for which CARDIOMETABOLIC RISK = Absence are used in the analysis phase.

The Scree Plot

The scree plot (Figure 21) is a graph of the eigenvalues against all the factors. The graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 2 and 3. Note also that factor 3 has an eigenvalue of less than 1, so only three factors have been retained. An eigenvalue is the standardized variance associate with a particular factor. The sum of the Eigen-values cannot exceed the number of items in the analysis, since each item contributes one to the sum of variances.

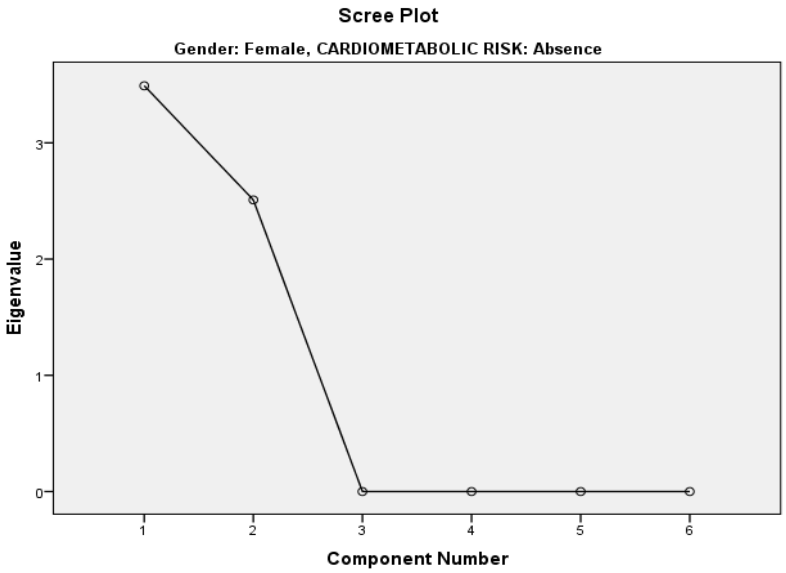

Figure 21 Scree Plot for Females in the Absence of Cardio-metabolic Risk.

Component Matrix

Error! Reference source not found. shows the loadings of the six variables on the two factors extracted. The higher the absolute value of the loading, the more the factor contributes to the variable. This analysis included all loadings for those that were more than 0.5. It is easily observed that the two factors can be identified as follows:

- Comprising of the variables; Diastolic Blood Pressure with a loading of 0.973, Waist Circumference with a loading of 0.941, Body Mass Index with a loading of -0.880, Glucose Fasting with a loading of 0.634 and Systolic Blood Pressure with a loading of -0.685. Medically these variables can be grouped to form the factors of Obesity, Sugar and Blood pressure;
- Comprising of the variables; Post Load Glucose -0.993, Glucose Fasting 0.774, Systolic Blood Pressure 0.728.

These three variables are medically grouped under the overall factors of sugar and partly as a result of blood pressure. It is understood here that the most important variable under this factor analysis that contributes to the influence of metabolic syndrome is identified to be Glucose. This means that as far as the research is concerned, sugar is the most important among all metabolic factors. This is followed quite closely by Obesity. This is contrary to earlier observed beliefs that blood pressure is more important among metabolic factors than other variables in the presence of Cardio metabolic Risk.

Table 34 Component matrix for females in the absence of cardio-metabolic syndrome.

	Component	
Variable	**1**	**2**
Systolic Blood Pressure	-.998	-.060
Diastolic Blood Pressure	.447	.895
Height	.910	.415
Body Mass Index	-.928	-.372
Glucose Fasting	-.189	.982
Waist Circumference	.341	.940
Post Load Glucose	.837	-.547

Rotated Component MatrixError! Reference source not found. presents the rotated component matrix for the component matrix analyzed above. The main advantage of rotation is to make the interpretation of the analysis easier. Looking at the above interpretation, it was not easy to create the three factors as separately as it was required. There is a sort of mix of the variables where you find that some factors were found under two or more components. The following table brings out a clearer component analysis. Of course, it must be understood that the higher the absolute value of the loading, the more the factor contributes to the variable. As noted above, applying the same logic here, the two components are as explained below:

- The first component is composed of three well defined variables; Waist Circumference with a loading of 0.978, Diastolic Blood Pressure with a loading of 0.948 and Glucose Fasting with a

loading of 0.946. These variables partly form the factors: Obesity, Blood pressure and Sugar respectively;

- The second component comprises Of Systolic Blood Pressure 0.980, Post Load Glucose -0.905, Body Mass Index 0.867. These variables area easily identified to partly respectively form the factors; Blood pressure, Sugar and Obesity.

Table 35 Rotated component matrix for females in the absence of cardio-metabolic risk.

	Component	
Variable	**1**	**2**
Systolic Blood Pressure	-.998	-.060
Diastolic Blood Pressure	.447	.895
Height	.910	.415
Body Mass Index	-.928	-.372
Glucose Fasting	-.189	.982
Waist Circumference	.341	.940
Post Load Glucose	.837	-.547

The plot below (see Figure 22) shows the items (variables) in the rotated factor space for women in the absence of cardio-metabolic syndrome. This was an SPSS output of the research data analysis. According to this figure, two factors were extracted (Table 35). The logic behind the factor formation is the high correlation of the variables with the respective factors. It is noted that the clustering of Height, Body Mass Index and Post load glucose formed the first factor "Obesity" or "Overweight" (F1). The second was a combination of two factors "Blood Glucose Metabolism Disordering" (F2) was formed by the clustering of Glucose fasting and Post Load Glucose and "Blood Pressure" formed by a clustering of Diastolic blood Pressure. The loading observed in table 5 were the loadings which were greater than 0.5 in absolute value.

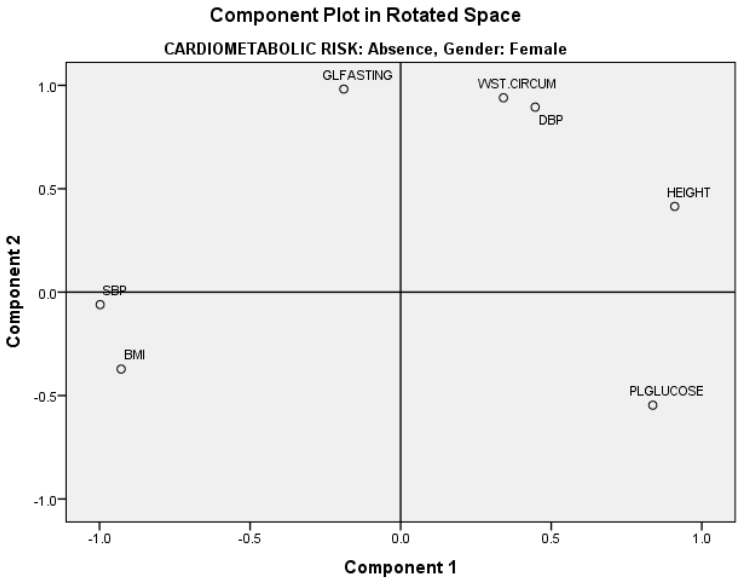

Figure 22 Component Plot in Rotated space for Females in the Absence of Cardio- metabolic Risk.

5.2.8. Findings in Males with Cardio-Metabolic Risk in Rural Setup

Descriptive Statistics

The descriptive output for the rural area and in the presence of cardio metabolic risk shows that among all the variables included in the study (**Error! Reference source not found.**). It is clear that the most important among them that influence the emergency of cardio-metabolic risk are; Post load glucose with a mean of 156.74 and a standard deviation of 78.044, Systolic blood pressure with a mean of 124.26 and a standard deviation of 23.034, Glucose Fasting with a mean of 91.67 and a standard deviation of 20.288, Diastolic blood pressure with a mean of 79.81 and a standard deviation of 17.017. The other variables have comparatively smaller means.

Table 36 Descriptive statistics for males in presence of cardio-metabolic syndrome in rural setups.

	Mean	Std. Deviation	Analysis N
Systolic Blood Pressure	124.26	23.034	54
Diastolic Blood Pressure	79.81	17.017	54
Height	1.6676	.11156	54
Body Mass Index	21.831	5.1851	54
Glucose Fasting	21.831	20.288	54
Waist Circumference	75.46	13.871	54
Post Load Glucose	156.74	78.044	54

Correlation Matrix for Males in Rural Setups with Metabolic Syndrome

Table 37 shows a matrix of linear correlations between some observed variables in the dataset. Like before, it was observed that since so many of the correlations were very small, this would have been a matrix in which all of the diagonal elements are 1 whereas all off diagonal elements are 0. However, this assumption was rejected by Bartlet's test whose p-value was 0.00, which meant that the off-diagonal elements of the correlation matrix were not all zero and in fact none was 0.00. It was an additional observation that the magnitude of the correlations ruled out the possibility of multicollinearity (high correlations between variables).

Also noted from Table 37 was that some pairs of variables had significant correlations; Body Mass Index vs Waist Circumference having a negative but significant correlation of -.666 ($P<0.01$), Body Mass Index vs Height with a negative but significant correlation of -0.356 and a p-value of less than 0.05, Waist Circumference vs Systolic Blood Pressure with a negative but significant correlation of -0.397 (p-value less than 0.05) have an indirect link with organizational ethical relatedness with the cardio-metabolic disease symptoms and factors that influence individual attack by metabolic syndrome factors. These two signify an indirect link with the factors of cardio-metabolic diseases. Glucose Fasting vs Post Load Glucose with a significant positive correlation and a (P-value < 0.05). The significant positive correlation of 0.480 has a direct and combined link with the existence of metabolic factors in the body. The p-value was less than 0.05. Body Mass Index vs Weight with a p-value of ($P<0.05$).

These significant correlations (with p-values less than 0.05) had a direct impact on the existence of Cardio-metabolic syndrome factors. The significant positive correlations demonstrate a high degree of association between the pairs which signify a direct internal organizational derivation of metabolic factors in an individual. In addition, waist circumference and glucose fasting with a significant but positive correlation of 0.859 (p-values less than 0.001), which shows a high level of association between the two variables. This high but significant association shows a direct link with cardio-metabolic factors which are strongly related to one another. The significant positive correlations show that the individual variables forming the pairs of strongly associated variables have equally significant associations with the underlying cardio-metabolic condition of an individual. The uncorrelated pairs of variables have little or no combined effect on the existence of cardio-metabolic syndrome factors, either directly or indirectly.

Table 37 Correlations "males in the presence" of cardio metabolic Risk in a rural setup.

Variable	Systolic BP	Diastolic BP	Height	Body MI	Glucose FA	Waist Circ.	Post Load Glucose
Systolic Blood Pressure	1	.630	.081	.076	.111	.086	-.218
Diastolic Blood Pressure	.630	1	.035	.007	.012	.150	-.35
Height	.081	.035	1	-.308	-.035	.200	-.075
Body Mass Index	.076	.007	-.308	1	.006	.021	-.134
Glucose Fasting	.111	.012	-.035	.006	1	-.089.584	
Waist Circumference	.086	.150	.200	.021	-.089	1	.052
Post Load Glucose	-.218	-.350	-.075	-.134	.584	.052	1

Communalities

Considering table 38, it is important to understand the amount of variance in the variables that has been accounted for by the extracted factors. Of course the initial accounting gives the 100% of the variance accounted for by the factors. This is why there is a string of 1s under the heading "Initial". From the output, it is understood that among the factors extracted and the corresponding variables, 89.8% of the variance in Waist circumference has been accounted for while 77.2% of the variance in Post load glucose has been accounted for. It is also noted that 82.1% of the variance in Body mass index has been accounted for. Other variances accounted for in other variables have been listed in the following table of communalities. The 89.8% of the variance accounted for meant that 89.8% of the proportion of variance in Waist Circumference was explained by the three factors with a total variance explained of 63.794%. Explained variation here measures the proportion to which a mathematical model accounts for the variation (dispersion) of a given data set. Quantified we say the amount of variance accounted for. 82.1% of the proportion of variance in Body Mass Index was explained by the three factors extracted. On the lower performance level, 23.9% of the proportion in Systolic Blood Pressure was explained by the three factors. The mathematical model performed better for some variables that it did for other variables. The worst was for the variable Systolic Blood Pressure followed by Glucose Fasting whose proportion of variation had was only explained 23.9% by the model.

Table 38 Communalities for males in the presence of cardio-metabolic risk in rural setups.

	Initial	Extraction
Systolic Blood Pressure	1.000	.239
Diastolic Blood Pressure	1.000	.713
Post Load Glucose	1.000	.772
Body Mass Index	1.000	.821
Glucose Fasting	1.000	.688
Waist Circumference	1.000	.898

Total Variance Explained

The next results (table 39) item shows all the factors extractable from the analysis along with their eigenvalues, the percent of variance attributable to each factor, and the cumulative variance of the factor and the previous factors. Notice that the first factor accounts for 26.759% of the variance, the second 20.370% and the third 16.665%. All the remaining factors are not at all significant. The principal component factor analysis has shown that in the rural setup of the target area in the presence of metabolic risk factors, three uncorrelated factors explained 63.794% of the observed variance. This has been further confirmed by the scree plot which has clearly distinguished the three factors based on their Eigen values. The corresponding Eigen values for the three factors are respectively: 2.299, 1.292 and 1.070. Principal Component Factor Analysis revealed three uncorrelated factors that cumulatively explained 63.794% of the observed variance in the absence of Cardio-metabolic risk. The number of those three factors was determined by the scree plot according to Eigen values and the total variance explained. These two factors could be identified as Blood Glucose Metabolism Disordering (Factor 1; % of variance=26.759), obesity (Factor 2; % of variance=20.370) and Blood Pressure (Factor 3; % of variance = 16.665%). (See table & scree plot below).

Table 39 Total variance explained for males in rural setups.

Initial Eigenvalues				**Extraction Sums of Squared Loadings**			**Rotation Sums of Squared Loadings**		
Component	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**
1	1.880	26.853	26.853	1.880	26.853	26.853	1.873	26.759	26.759
2	1.532	21.892	48.746	1.532	21.892	48.746	1.426	20.370	47.129
3	1.053	15.049	63.749	1.053	15.049	63.794	1.167	16.665	63.794
4	.982	14.022	77.816						
5	.722	10.312	88.128						
6	.568	8.121	96.249						
7	.263	3.751	100.00						

The Scree Plot

Error! Reference source not found. describes the scree plot which is a graph of the Eigen values against all the factors. This graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 3 and 4. Note also that factor 4 has an Eigen value of less than 1, so only three factors have been retained.

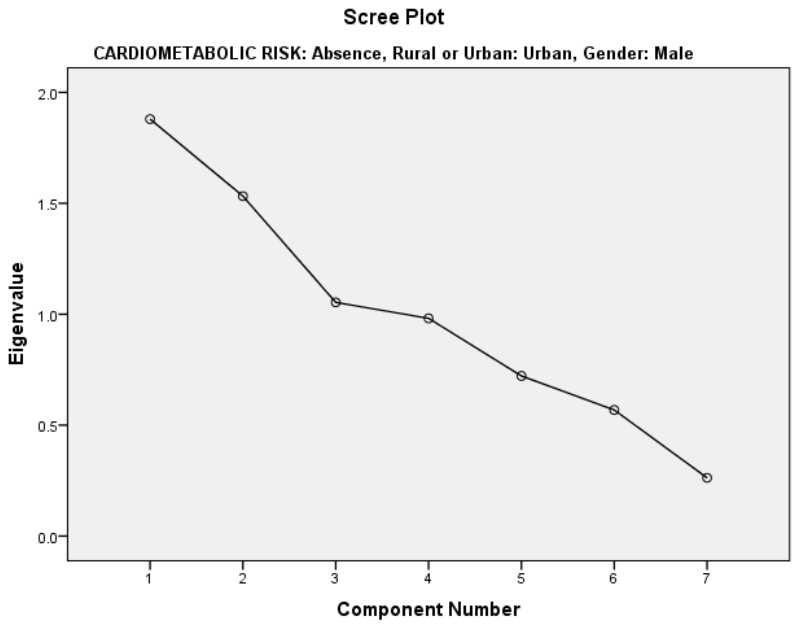

Figure 23 Scree Plot in the Presence of Cardio-metabolic Risk for Males in Rural Setups.

Component Matrix

Error! Reference source not found. shows the loadings of the six variables on the three factors extracted. The higher the absolute value of the loading, the more the factor contributes to the variable. This analysis included all loadings even those that were less than 0.5. It is easily observed that the three factors can be identified as follows:

- Component 1 comprising of the variables; Glucose fasting with a loading of POST Load Glucose with a loading of -0.814, Glucose Fasting with a loading of -0.793, Diastolic Blood Pressure with a loading of 0.748 and Systolic Blood Pressure with a loading of 0.639. Medically these four variables can be grouped to form the two factors of *Sugar (glucose) and Blood pressure respectively;*

- Component 2 comprising of the variables; Body Mass Index with a loading of -0.752 and Systolic Blood Pressure with a loading of 0.534. These two variables partly form the factors of obesity and blood pressure;
- Component 3 comprising of Waist Circumference with a loading of 0.899 and Body Mass Index with a loading of -0.434. These two from the overall factor of *Obesity*.

It is understood here that the most important variable for this factor analysis that contributes to the influence of metabolic syndrome is identified to be Glucose and Blood pressure. This means that as far as the research is concerned, sugar and blood pressure are the most important among all metabolic factors. This is followed quite closely by Obesity. This is contrary to earlier observed beliefs that blood pressure is more important among metabolic factors than other variables among males in the presence of Cardio metabolic Risk in a rural setup.

Table 40 Component Matrix for Males in the Presence of Cardio-metabolic Risk in Rural Setups.

	Component		
Variable	**1**	**2**	**3**
Systolic Blood Pressure	.639	.534	-.247
Diastolic Blood Pressure	.748	.410	.090
Body Mass Index	.122	-.752	-.434
Waist Circumference	.155	-.256	.899
Glucose Fasting	-.793	.316	-.048
Post load Glucose	-.814	.327	.042

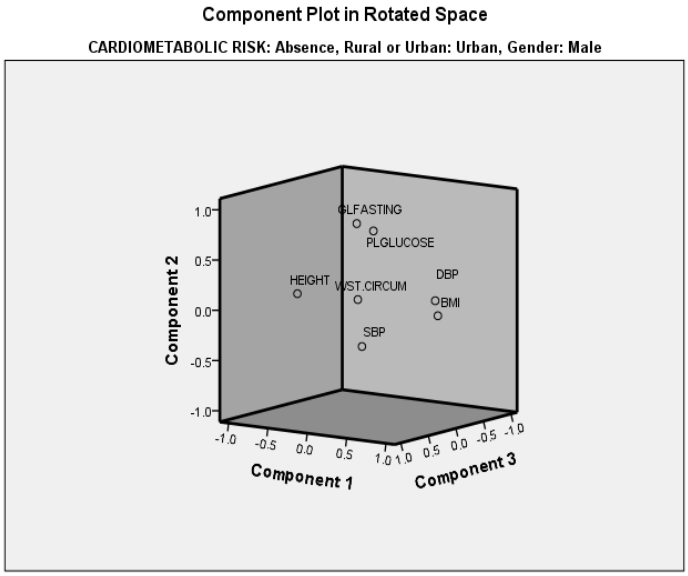

Figure 24 Component Plot In Rotated Space in the Presence of Cardio- metabolic Risk.

Rotated Component Matrix

Error! Reference source not found. above is the Rotated Component Matrix for Males in the Presence of Cardio-metabolic Risk in a Rural Setup. The loadings in this table make it easier to see the combinations of the variables which easily lead to the identification of the factors. Looking at the table, we can see that Systolic blood pressure, and diastolic blood pressure are substantially loaded on Factor (Component) 1, while Body mass index, glucose fasting and post load glucose are substantially loaded on Factor 2 whereas only waist circumference is substantially loaded on Factor (Component) 3. These factors can be used as variables for further analysis.

5.2.9. Critical Predictions of the Exploratory Factor Analysis results

This section gives a critical discussion of the results observed under the preceding sections. The researcher noted here that there seemed to be a relationship between the presence of Cardio-metabolic Syndrome and the setup of collection of the data. There were different setups identified for the comparative analysis of these study. These were: In the general population, for males only, for females only, under the rural or urban setup, and under a combination of any setup with another.

This discussion will be limited to the setups discussed below. The researcher believed that demonstration of the outcome of a result based on given limited setups, was as important as a

demonstration of all available setups. The fact is not necessarily to give full proof, but rather, to demonstrate the evidence of the existence of a hypothetical finding using part or the whole collected data. For this discussion, the researcher will include information on the following setups:

- Under the general population;
- In the presence of Cardio-metabolic Syndrome;
- In the absence of Cardio-metabolic Syndrome
- Among the males;
- Among the females;
- Among males in the presence of Cardio-metabolic Syndrome;
- Among females in the absence of Cardio-metabolic Syndrome;
- Females in the absence of cardio-metabolic Syndrome.

According to results obtained, the use of Principal Component Analysis produced three components in the general population which were identified as; Factor 1 formed by Systolic Blood Pressure and Diastolic Blood Pressure with absolute loadings of 0.859 and 0.856 respectively. Factor 2 was constituted by the combination of Height with a loading of 0.878 and Body Mass Index with a loading of 0.889. Finally, factor 3 was composed of Glucose Fasting and Post Load Glucose. Under the same setup, the estimated averages for Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index were respectively; 117.47, 70.39 and 29.981. The component plot in rotated space showed well-constructed combined variables forming the observed components demonstrating the consistent structure of the formation of the three components under the general population.

On the other hand, in the absence of Cardio-metabolic Syndrome, Glucose Fasting and Post Load Glucose formed factor 1 while Body Mass Index and Waist Circumference formed the second factor. Similarly the respective estimated averages for Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index were respectively; 113.41, 66.56 and 24.677. The researcher observed significant changes which were notably; a sharp and significant drop in the three averages of Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index as compared to the general population readings. This indicated that Cardio-metabolic Syndrome factors were directly influenced by setup from which the data were drawn. The more favorable condition being in the absence of Cardio-metabolic Syndrome.

In the presence of the Cardio-metabolic Syndrome of the general population, three factors were extracted which were identified as follows: Glucose Fasting and Post Load Glucose created factor 1, with respective loadings of 0.906 and 0.894. Systolic Blood Pressure and Diastolic Blood Pressure led to Factor 2 and Body Mass Index formed factor 3. Here, the estimated averages of Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index were respectively; 127.60, 80.00 and 43.40.

This brings to fore the earlier observed finding that the presence of Cardio-metabolic Syndrome influenced the elevation of Cardio-metabolic factors such as higher than normal blood pressure leading to hypertension. This signals the need for earlier warning to identified individuals of a possible attack of opportunistic Cardio-metabolic Risk factors. Such warning should be followed by medical attention to prevent a full blown condition.

When males in the presence of metabolic syndrome were isolated from the rest, three factors were extracted by principal component analysis. The estimated averages of Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index were respectively; 127.62, 79.95 and 43.367. The three factor-extractions were combinations as follows:

- Factor 1 (Height & Body Mass Index);
- Factor 2 (Glucose Fasting & Post Load Glucose); and
- Factor 3 (Systolic Blood Pressure). All the state variables had heavy absolute loadings on the extracted factors.

At this point, this researcher claims with authority that given the heavy absolute loadings on the three extracted factors and the significant high loadings on the factors by the variables, the existence of a very strong relationship between Metabolic Syndrome and the presence or absence of Cardio-metabolic Risk factors is a fact well-established.

To crown it all, the researcher considered results of males in the absence of Cardio-metabolic Risk factors. The estimated averages of Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index were respectively; 113.33, 66.59 and 24.588. Under this setup, the principal component analysis extracted only two factors. The two factors had the following information: Factor 1 (Height & Body Mass Index) and factor 2 (Glucose Fasting & Post Load Glucose). A total variance explained of 47.607% was displayed.

The last setup data was collected on female participants in the absence of metabolic syndrome. For this setup, two factors were extracted. The two factors had the following information: Factor 1 (Systolic Blood Pressure, Height, Body Mass Index & Post Load Glucose) and factor 2 (Diastolic Blood Pressure, Glucose Fasting & Waist Circumference). Here the estimated averages of Systolic Blood Pressure, Diastolic Blood Pressure and Body Mass Index were respectively; 120, 63.33 and 27.206. Under this setup, the Total Variance Explained was 100.00%. The component plot in rotated space did not demonstrate a clearly well-defined structure of component formation.

An immediate observation was the low average readings of the stated variable. This, in addition to other previous analyses, leaves no doubt about the predisposition and prevalence of Cardio-metabolic Syndrome factors. Simply put, the predisposition and prevalence were notably higher in males than in the female counterparts and further, that the vulnerability of males in the presence of Cardio-metabolic risk had a higher probability that it was among the females.

5.3. Confirmatory Factor Analysis in the General Population

This section presents analysis, interpretation and discussion on results obtained using Confirmatory Factor Analysis.

5.3.1. AMOS when Cardio-Metabolic Syndrome is an Exogenous Variable

The researcher started by computing ordinary multivariate analysis where two dependent variables were classified as being exogenous whereas other variables in the models were exogenous variables. Consequently, Cardio-metabolic Syndrome and Diabetes were exogenous. The multivariate analysis was performed by the use of SPSS AMOS version 23. The result obtained have been summarized below. AMOS (Analysis of Moment Structures) an SPSS package used when the variable (Cardio-metabolic Syndrome) was the Exogenous variable while SBP (Systolic Blood Pressure), WAIST.CIRCUM (Waist Circumference), Weight, Height, DBP (Diastolic Blood Pressure) and BMI (Body Mass Index) were endogenous variables.

5.3.2. SPSS AMOS Used For Multivariate Linear Regression

Error! Reference source not found. below shows the maximum likelihood estimates of a variance-covariance matrix used for CFA to estimate the variance–covariance matrix (Rencher, 2002).

Table 41 Confirmatory Factor Analysis prediction of the variance-covariance matrix.

$\lambda_{11}^2 + \psi$	$\lambda_{11}\lambda_{21}$	$\lambda_{11}\lambda_{31}$	$\lambda_{11}\ \phi_{12}\lambda$	$\lambda_{11}\ \phi_{12}\lambda$	$\lambda_{11}\phi_{12}\lambda$
$\lambda_{11}\lambda_{21}$	$\lambda_{21}^2 + \psi$	$\lambda_{21}\lambda_{31}$	$\lambda_{21}\phi_{12}\lambda$	$\lambda_{21}\phi_{12}\lambda$	$\lambda_{21}\phi_{12}\lambda$
$\lambda_{11}\lambda_{31}$	$\lambda_{21}\lambda_{31}$	$\lambda_{31}^2 + \psi$	$\lambda_{31}\phi_{12}\lambda$	$\lambda_{31}\phi_{12}\lambda$	$\lambda_{31}\phi_{12}\lambda$
$\lambda_{11}\ \phi_{12}\lambda$	$\lambda_{21}\phi_{12}\lambda$	$\lambda_{31}\phi_{12}\lambda$	$\lambda_{42}^2 + \psi$	$\lambda_{42}\lambda_{52}$	$\lambda_{42}\lambda_{62}$
$\lambda_{11}\ \phi_{12}\lambda$	$\lambda_{21}\phi_{12}\lambda$	$\lambda_{31}\phi_{12}\lambda$	$\lambda_{42}\lambda_{52}$	$\lambda_{52}^2 + \psi$	$\lambda_{42}\lambda_{62}$
$\lambda_{11}\phi_{12}\lambda$	$\lambda_{21}\phi_{12}\lambda$	$\lambda_{31}\phi_{12}\lambda$	$\lambda_{42}\lambda_{52}$	$\lambda_{52}\lambda_{62}$	$\lambda_{62}^2 + \psi$

Table 42 Regression weights and estimates: Cardio-metabolic Risk.

Dependent variable	Independent variable	Estimate	S.E.	C.R.	P
SBP	Glucose Fasting	.016	.007	2.466	.014
Weight	DBP	.012	.038	.324	.746
Waist Circumference	BMI	.015	.010	1.533	.125
Cardio-metabolic	SBP	-.002	.001	-3.353	***
Cardio-metabolic	Glucose Fasting	-.002	.000	-12.676	***
Cardio-metabolic	DBP	-.015	.001	-15.353	***
Cardio-metabolic	Weight	-.001	.001	-.734	.463
Cardio-metabolic	Height	-.009	.006	-1.653	.098
Cardio-metabolic	Waist Circumference	.000	.001	-.030	.976
Cardio-metabolic	BMI	.000	.000	-.943	.345

Table 43 Standardized Regression Weights (Group 1-Default model).

Dependent variable	Independent variable	Estimate
SBP	Glucose Fasting	.079
Weight	DBP	.010
Waist Circumference	BMI	.071
Cardio-metabolic	SBP	-.090
Cardio-metabolic	Glucose Fasting	-.341
Cardio-metabolic	DBP	-.411
Cardio-metabolic	Weight	-.020
Cardio-metabolic	Height	-.044
Cardio-metabolic	Waist Circumference	-.001
Cardio-metabolic	BMI	-.025

Table 44 Means: (Group 1 - Default model).

Variable	Estimate	S.E.	C.R.	P	Label
BMI	26.496	2.149	12.331	***	Par_15
Glucose Fasting	116.883	3.390	34.474	***	Par_16
DBP	71.630	.428	167.510	***	Par_17
Height	1.723	.072	23.997	***	Par_14

Table 45 Intercepts: (Group 1 - Default model).

	Estimate	**S.E..**	**C.R.**	**P**	**Label**
SBP	117.334	1.034	113.526	***	Par_12
Weight	64.9302	2.783	23.334	***	Par_13
Waist Circumference	79.627	.712	111.871	***	Par_19
Cardio-metabolic risk	3.111	.152	20.424	***	Par_18

Table 46 6 Variances: (Group 1 - Default model).

Variable	estimate	S.E.	C.R.	P	Label
Glucose Fasting	11189.508	507.290	22.057	***	Par_20
DBP	178.023	8.07122.058	22.057	***	Par_21
BMI	4506.474	203.998	22.091	***	Par_22
E7	202.304	13.317	15.191	***	Par_23
E6	467.546	21.3198	22.057	***	Par_24
E8	252.790	11.443	22.091	***	Par_25
Height	5.031	.228	22.091	***	Par_26
E5	.157	.007	22.071	***	Par_27

Table 47 Squared multiple Correlations: (Group number 1 - Default model) for Cardio-metabolic risk.

Variable	Estimate
Waist Circumference	.005
Weight	.000
SBP	.006
Cardio-metabolic Risk	.302

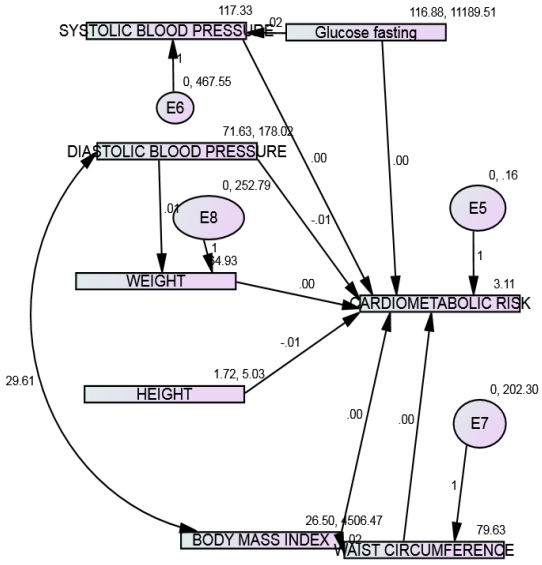

Figure 25 Multiple Linear Regression between the Cardio-metabolic Syndrome and endogenous variables.

Results and Discussion

Descriptive Statistics

Table 44 shows that the mean for BMI was 26.496 which was highly significant with a p-value of less than 0.001. Glucose Fasting had a mean of 116.883 and DBP (71.630), Height (1.723). All the observed means were highly significant. The significance in this case was due to the individual variable data as compared to other data collected elsewhere. The reason for this reasoning is that

the variables were completely independent of each other and had different measurement instruments. While 1.723 was significant as a mean for Height in meters, it was too insignificant as a mean for DBP measured in completely different units. This means that all the variables had significant and meaningful means. The dependent variable here is Cardio-metabolic risk. The larger the intercept for a given variable, the more it contributes to cardio-metabolic risk.

The Intercepts

Table 45 shows the intercepts and their corresponding p-values for some model variables as stated here: SBP (117.334, ***), Weight (64.9302, ***), Waist Circumference (79.627, ***), Cardio-metabolic risk (3.111, ***). It is noted that for reasons stated above, the intercepts are larger for some variables than others. The reason is that the measurements are based on different measuring instruments and the units of measurement are also different. When a variable has a larger intercept than another, a direct comparison will not carry sense. The research finds that the larger the intercept, the more it contributes to cardio-metabolic risk.

Estimates

Results in Table 42 gives the dependent variable and the independent variable and their corresponding p-value of regression for the following combination of variables in the model fit. SBP--Glucose Fasting (.014), Weight—DBP (.746), Waist Circumference—BMI (.125), Cardio-metabolic—SBP (***), Cardio-metabolic--Glucose Fasting (***), Cardio-metabolic—DBP (***), Cardio-metabolic—Weight (.463), Cardio-metabolic—Height (.098), Cardio-metabolic--Waist Circumference (.976), Cardio-metabolic—BMI (.345) Byrne (2001), Hoffmann (2008).

Table 42 shows that the estimation of weight using DBP is not significant, SBP using glucose fasting is significant, cardio-metabolic risk using SBP is significant, cardio-metabolic risk using glucose fasting is significant, cardio-metabolic risk using SBP is significant, cardio-metabolic risk using DBP is significant, cardio-metabolic risk using weight is not significant, Cardio-metabolic using waist Circumference is not significant, Cardio-metabolic using BMI is not significant. It is understood that using this model, the estimation of cardio-metabolic risk is fairly significant in most cases. It is evident that the estimation of cardio-metabolic risk is quite successful using the current model.

Variable Variances Estimation and P-values

Table 46 of group 1, the default model gave the following variables against their variances and the respective p-values. Glucose Fasting (11189.508, ***), DBP (178.023, ***), BMI (4506..474, ***), E7 (202.304, ***), E6 (467.546, ***), E8 (252.790, ***), Height (5.031, ***), E5 (.157, ***). The highly significant variances define variables with large variances and which are not equated to "0". All the above variables have significantly very large variances which means that the measurements for data collection provide wide ranges of data from very small to very large values.

AMOS when All Diabetics is Exogenous Variable

We have three types of variables in our analysis. Which are, observed endogenous, observed exogenous, and unobserved exogenous variables. The table (see Table 48) below categorized according to the three type of variable (observed endogenous, observed exogenous and unobserved exogenous variables).

Table 48 Types of variables to be used in the analysis.

Observed, Endogenous Variables	Observed, Exogenous Variables	Unobserved, Exogenous Variables
SBP WAIST.CIRCUM DIABETES DBP WEIGHT HEIGHT	GLFASTING BMI	E6 E7 E8 E9 E10 E11

Table 49 Regression Weights and Estimates: All Diabetics.

Dependent variable	**Independent variable**	**Estimate**	**S.E.**	**C.R.**	**P**	**Label**
SBP	Glucose Fasting	.016	.007	2.495	.013	Par_1
Waist Circumference	BMI	.015	.010	1.533	.125	Par_2
DBP	Weight	-.001	.027	-.024	.981	Par_10
DBP	Height	-.191	.189	-1.008	.31.	Par_11
DBP	BMI	.006	.006	1.009	.313	Par_12
SBP	BMI	-.008	.010	-.745	.456	Par_13
DBP	Glucose Fasting	.016	.004	3.950	***	Par_14
Diabetes	Glucose Fasting	.002	.000	19.282	***	Par_3
Diabetes	SBP	.001	.001	1.557	.120	ParP_4
Diabetes	DBP	.004	.001	4.386	***	Par_5
Diabetes	Weight	.000	.001	.369	.712	Par_6
Diabetes	Height	.008	.005	1.553	.120	Par_7
Diabetes	BMI	.000	.000	1.722	.085	Par_8
Diabetes	Waist Circumference	.001	.001	.446	.655	Par_9

Table 50: Standardized Regression Weights: (Group number 1- Default model).

Dependent variable	**Independent variable**	
SBP	Glucose Fasting	.080
Waist Circumference	BMI	.071
DBP	Weight	-.001

DBP	Height	-.032
DBP	BMI	.032
SBP	BMI	-.024
DBP	Glucose Fasting	.126
Diabetics	Glucose Fasting	.519
Diabetics	SBP	.042
Diabetics	DBP	.118
Diabetics	Weight	.010
Diabetics	Height	.041
Diabetics	BMI	.046
Diabetics	Waist Circumference	.017

Table 51 Means: (Group 1 - Default model).

	Estimate	**SE..**	**C.R.**	**P**	**Label**
BMI	26.496	2.149	12.331	***	Par_18
Glucose Fasting	117.057	3.393	34.504	***	Par_19

Table 52 Intercepts: (Group number 1- Default model).

	Estimate	**S.E.**	**C.R.**	**P**	**Label**
Weight	65.816	.5096	129.316	***	Par_16
Height	1.723	.072	23.997	***	Par_17
SBP	117.514	1.069	109.925	***	Par_15
DBP	69.968	1.901	36.811	***	Par_20
WAIST Circumference	79.617	.712	111.870	***	Par_21
Diabetes	-.449	.141	-3.191	.001	Par

Table 53 Variances: (Group 1 - Default model).

	Estimate	**S.E.**	**C.R.**	**P**	**Label**
E8	252.818	11.445	22.091	***	Par_23
E9	5.031	.228	22.091	***	Par_24
Glucose Fasting	11208.537	508.101	22.060	***	Par_25
BMI	4506.474	203.998	22.091	***	Par_26
E6	467.216	21.183	22.057	***	Par_27
E7	202.301	13.317	15.191	***	Par_28
E10	174.863	7.928	22.056	***	Par_29
E11	.134	.006	22.062	***	Par_30

Table 54 Squared multiple Regressions: (Group 1-Default model).

	Estimate
Height	.000
Weight	.000
DBP	.018
Waist Circumference	.005
SBP	.007
Diabetes	.309

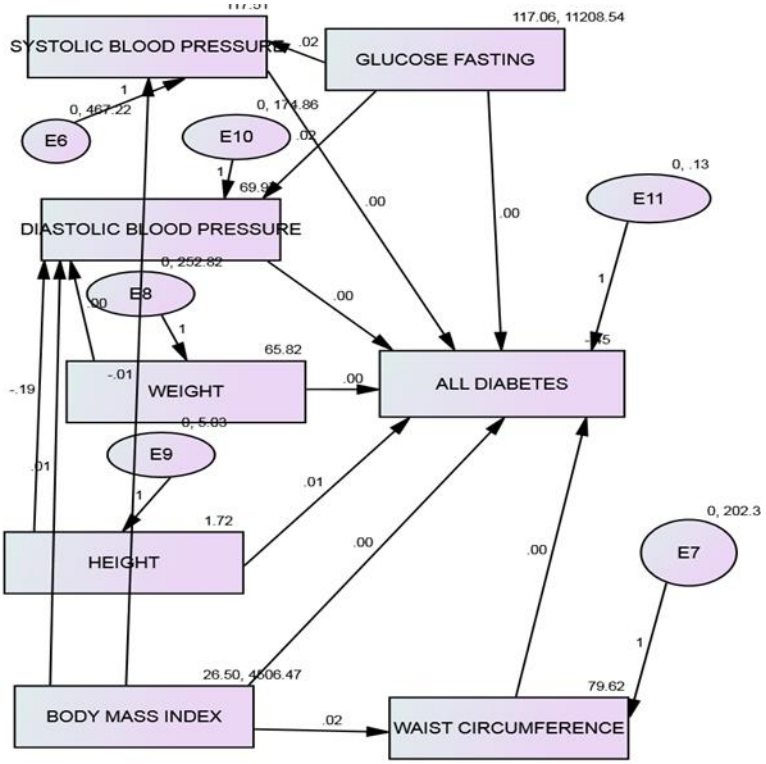

Figure 26 Multiple Linear Regression between All Diabetes and endogenous manifest variables.

Results and Discussion

Figure 26 indicates that the relationship between the independent variables and the dependent variable (All Diabetes) is significant for some but not for others. It is noted that All diabetics is strongly influenced by glucose fasting with a regression coefficient weight of 0.002 and a p-value of less than 0.001, while diastolic blood pressure has a significant influence on all diabetics with a regression weight of 0.004 and a p-value less than 0.001 (Hair et al., 2006). These estimates are further supported by those in (Table 53 and Table 54). The analysis also shows that the two variables have both direct and indirect effects on all diabetics. The variables which do not influence all diabetics are indicated below with their respective regression coefficients and the corresponding p-values: SBP with a regression weight of .001 and a p-value of .120, Weight with a regression weight of .000 and a p-value of .712, Height with a regression weight of .008 and a p-value of .120, BMI with a regression weight of .000 and a p-value of .085, Waist Circumference with a regression weight of .001 and a p-value of .655 (Muller et al. 2005), (Baron and Kenny, 1986). These variables which have no direct influence on the symptomatic behaviour of diabetes may

have indirect influences of diabetes. The specific diabetes in question is Type 2 diabetes mellitus whose prevalence is high in Africa.

Table 54 also suggests that most of the independent variables in the model are not directly associated with one another. In fact the only ones which have a significant association are Glucose Fasting and systolic blood pressure with a p-value of .013, Waist Circumference---BMI with a p-value of .125, DBP---Weight with a p-value of .981, DBP---Height with a p-value of .310, DBP---BMI with a p-value of .313 and SBP---BMI with a p-value of .456 (Schueth, 2003).Table 56 shows estimated means for those variables which highly influenced the dependent variable. The reader can remember that in the foregoing discussion, BMI and glucose fasting both independently influence the manifestation of diabetes. Their estimated means were respectively found to be as stated below: BMI with an estimated mean of 26.496 and a highly significant p-value less than 0.001 denoted by (***) and Glucose Fasting with an estimated mean of 117.057 and a highly significant p-value far less than 0.001 and also denoted by (***).

In addition to the above discussion, table 57 provides intercepts for the model in this discussion which show that all the intercepts are significant. Variables with their respective intercepts are listed here with the corresponding p-values and the level of significance. Weight (65.816, ***), Height (1.723, ***), SBP (117.514, ***), DBP (69.968, ***) and Waist Circumference (79.617, ***) and Diabetes (-.449, 0.001) Shafer W.E., (2008). It was commented earlier that there was a possibility of some independent variables not having direct influence on the dependent variable but have a high influence through some mediating variable. A part from the above information, table 58 for the group 1 (the default model) gives variances for the recorded variables in the model. According to the analysis, this research observed that different variables had different variances. The variables, the respective estimated variance and the p-value are listed as follows: E8 (252.818, ***), E9 (5.031, ***), Glucose Fasting (11208.537, ***), BMI (4506.474, ***), E6 (467.216, ***), E7 (202.301, ***), E10 (174.863, ***), E11 (.134, ***). There is need to understand the levels of variance for each of the variables related with the response variable in this study. The degree of variability of the independent variables will guide in the understanding different measurements of diabetes in this case. A collection of stable independent variables will signify a more tolerable situation. Thus it is important to understand the variance structure of the variables involved in the estimation and modelling of diabetes for a successful research. Table 59 provides estimated variances for independent variables in the model. It is estimated that the predictors of height

explain 0.000 percent of its variance. In other words, the error variance of height is approximately 100 percent of the variance of height itself. It is estimated that the predictors of weight explain 0.000 percent of its variance. In other words, the error variance of weight is approximately 100 percent of the variance of weight itself, It is also estimated that the predictors of DBP explain 1.8 percent of its variance. In other words, the error variance of DBP is approximately 98.2 percent of the variance of height itself. Further, it is estimated that the predictors of diabetes explain 30.9 percent of its variance. In other words, the error variance of diabetes is approximately 69.1 percent of the variance of diabetes itself. It is estimated that the predictors of SBP explain 0.700 percent of its variance. In other words, the error variance of SBP is approximately 99.30 percent of the variance of SBP itself.

SPSS AMOS Confirmatory Factor Analysis in the general population

A layout of the guideline on followed while performing Confirmatory Factor Analysis is explained in chapter 3. Results that follow in the following section is Confirmatory Factor analysis output among the general population.

Table 55 Regression Weights and Estimates: Confirmatory Factor Analysis in the general population.

Dependent variable	Independent variable	Estimate	S.E.	C.R.	P	Label
SBP	F1	13.535	1.832	7.388	***	b1
DBP	F1	12.219	1.589	7.691	***	B2
Glucose Fasting	F2	96.737	4.354	22.218	***	B3
Post Load Glucose	F2	220.828	9.155	24.122	***	B4
BMI	F3	65.607	1.558	42.109	***	B6
Waist Circumference	F3	1.038	.678	1.532	.126	B7

Table 56 Standardized Regression Estimates: (Group 1- Default model).

Dependent variable	Independent variable	Estimate
SBP	F1	.624
DBP	F1	.915
Glucose Fasting	F2	.914
Post Load Glucose	F2	1.092
BMI	F3	.977
Waist Circumference	F3	.073

Table 57 Intercepts: (Group number 1 - Default model).

Dependent variable	Estimate	S.E.	C.R.	P
SBP	119.234	.696	171.427	***
DBP	71.624	.428	167.386	***
Glucose Fasting	117.109	3.388	34.562	***
Post Load Glucose	199.390	6.697	29.773	***
BMI	26.496	2.149	12.331	***
Waist Circumference	80.028	.663	120.731	***

Table 58 Covariance: (Group number 1-Default model).

		Estimate	S.E.	C.R.	P
F1	F2	.145	.037	3.891	***
F2	F3	.019	.029	.664	
F3	F1	.029	.036	.802	

Table 59 Correlations: (Group 1 - Default model).

		Estimate
F1	F2	.145
F2	F3	.019
F3	F1	.029

Table 60 Variances: (Group 1-Default model).

	Estimate	S.E.	C.R.	P	Label
F1	1				
F2	1				
F3	1				
e11	287.540	48.413	5.939	***	e1
e21	28.854	38.026	.759	.448	e2
e31	1845.155	682.873	2.702	.007	e3
e23	-5525.242	3538.400	-1.562	.118	e4
e13	202.255	13.314	15.191	***	e6
e24	202.255	13.314	15.191	***	e6

Table 61 Squared Multiple Correlations: (Group number 1-Default model).

	Estimate
Waist Circumference	.005
BMI	.955
Post Load Glucose	1.128
Glucose Fasting	.835
DBP	.838
SBP	.389

Table 62 Factor Score Weights (Group number 1-Default model).

	WC	BMI	PLG	FG	DBP	SBP
F1	.000	.000	.015	-.019	-.003	.000
F2	.000	.000	.000	.000	.062	.007
F3	.000	.015	.000	.000	.000	.000

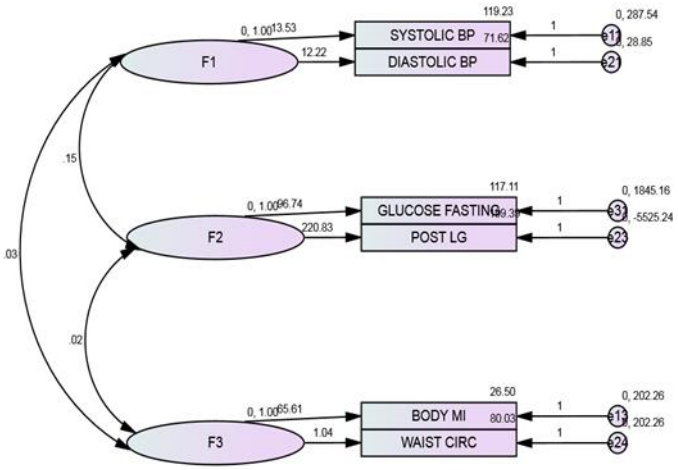

Figure 27 Confirmatory Factor Analysis in the general population using AMOS.

Results and Discussion

Descriptive Statistics

Table 62 gives the observed intercepts for the endogenous variables in the CFA Amos model of the setup of analysis. The intercepts for the different variables in the model were state as follows:

SBP (119.234, ***), DBP (71.624,***), Glucose Fasting (117.109, ***), Post Load Glucose (199.390, ***), BMI (26.496, ***) and Waist Circumference (80.028, ***). From this list , one notes that post load glucose has the highest intercept showing that on average, post load glucose had generally a higher starting point that others and that in its regression with its factor (F2), it is quite far from the origin. Another important regression variable is SBP whose intercept was 119.234. The third highest intercept corresponds to glucose fasting with a value of 117.109. It is noted that from the regression point of view, the intercepts are very important as they are all highly significant (Hellsten and Mallin , 2006).

Figure 27 suggests that based on the loadings of the path diagram, Systolic Blood Pressure has a higher influence on factor 1 (F1) than Diastolic Blood Pressure. Remember that factor 1 (F1) is Blood Pressure. From the medical point of view, this research and analysis of the data reveals that elevated Systolic Blood Pressure is seen to be more important to blood pressure construct than diastolic blood pressure. This is from the comparative point of view and it does not mean that

diastolic blood pressure has no significance. This information is strongly supported by the highly significant p-value (less than 0.001) and only indicated as (***) table 62 Byrne (2001).

Again, from Figure 27, the researcher noted that Post Load Glucose with a loading of 220.83 had a higher influence on factor 2 (F2) than Glucose Fasting. Factor 2 is glucose. Remember that glucose or sugar level in the body has a big influence on the condition of diabetes or diabetic condition in the human body. Glucose fasting had a loading of 96.74. In the medical profession, there has been a strongly worded controversy over the use of which of the two criteria to use for testing the emergency and level of diabetes in humans.

Finally, this research has revealed that with regard to the third factor (F2), BMI (Body Mass Index) has a higher influence on this factor than waist circumference. The higher loading of 65.607 on the BMI variable for factor 2 (F2) with a p-value less than 0.001 table 60 and figure 27 Hellsten (2006).

A further discussion of the analysis and interpretation considers the covariances between pairs of the factors. The factors and their corresponding covariances are listed as; F1-F2 (.145), F2-F3(.019), F3-F1 (.029). From table 59, only the first pair has a significant variance. This means that the presence of factor 1 has an influence on the manifestation of factor 2 and vice-versa. From plain words, the presence of blood pressure has a high influence on the elevation/emergency of glucose (diabetes). That is to say that high levels of both systolic and diastolic blood pressure have significant effects on the prevalence of diabetes, table 59. To sum it up, it means that the presence of high blood pressure is likely to signal the presence of diabetes. The combinations of F2-F3(.019), F3-F1 (.029) were not significant. Table 64 shows a table of correlation coefficients which have a similar approach due to their relatedness. Once more, factor 1 and factor 2 are related in such a way that the presence of blood pressure signals the presence of diabetes. These results show that the two are likely to manifest together though not necessarily at the same time. This means that having noticed the presence of one, precautions should be taken against the other Byrne (2001), Hoffmann (2008).

Table 65 shows the variances and the corresponding p-values of independent variables in the model, which are; e11 (287.540,***), e21(28.854, .448), e31 (18450.155, .007), e23 (-5525.242, .118), e13 (202.255, ***), e24 (202.255, ***). The variances of the independent variables have strong effects on the distribution of the dependent variable. The more variance the variable has,

the more it has effects on the measurements of the dependent variable. Accordingly, e31, e23, e11 and e13 and e24 have serious consequences on the data stability of the factors.

It is estimated that the predictors of Waist Circumference explain 0.5 percent of its variance. In other words, the error variance of Waist Circumference is approximately 99.5 percent of the variance of Waist Circumference itself. It is estimated that the predictors of BMI explain 0.5 percent of its variance. In other words, the error variance of BMI is approximately 99.5 percent of the variance of BMI itself. It is estimated that the predictors of Glucose Fasting explain 83.5 percent of its variance. In other words, the error variance of Glucose Fasting is approximately 99.5 percent of the variance of Glucose Fasting itself.

5.3.3. Factor Analysis for Data Split According to the Absence of Metabolic Syndrome

This section presents Confirmatory Factor Analysis output in the absence of Metabolic Syndrome together with the required interpretation.

Table 63 Definitions of variables names used in some tables.

DEFINITIONS OF VARIABLES USED UNDER THE FOLLOWING TABLES	
Variables name used	Meaning
VAR00059	Systolic Blood Pressure
VAR00060	Diastolic Blood Pressure
VAR00063	Body Mass Index
VAR00064	Fasting Glucose
VAR00065	Waist Circumference
VAR00076	Post Load Glucose

Table 64 Regression Weights and Estimates: Confirmatory Factor Analysis (In absence of cardio-metabolic risk).

Variable	**Factor**	**Estimate**	**S.E.**	**C.R.**	**P**	**Label**
VAR00059	F1	5.377	1.923	2.796	.005	b1
VAR00060	F1	1.779	.692	2.572	.010	b2
VAR00064	F2	7.312	1.586	4.610	***	b3
VAR00076	F2	15.148	3.141	4.823	***	b4
VAR00063	F3	1.584	.531	2.983	.003	b6
VAR00065	F3	11.202	.719	15.583	***	b7

Table 65 Standardized Regression Weights: Confirmatory Factor Analysis (In the absence of cardio-metabolic risk).

Variable	**Factor**	**Estimate**
VAR00059	F1	.433
VAR00060	F1	.231
VAR00064	F2	.517
VAR00076	F2	.844
VAR00063	F3	.209
VAR00065	F3	.832

Table 66 Intercepts for Confirmatory Factor Analysis in the absence of cardio-metabolic risk.

	Estimate	S.E.	C.R.	P	Label
VAR00059	113.104	.518	218.443	***	
VAR00060	66.212	.321	206.258	***	
VAR00064	82.668	.590	140.028	***	
VAR00076	123.205	.749	164.407	***	
VAR00063	24.513	.318	77.003	***	
VAR00065	79.852	.815	97.979	***	

Table 67 Covariances for Confirmatory Factor Analysis in the absence of cardio-metabolic risk.

		Estimate	S.E.	C.R.	P	Label
F1	F2	-.342	.140	-2.438	.015	
F2	F3	.121	.084	1.446	.148	
F3	F1	.166	.154	1.076	.282	

Table 68 Correlations for confirmatory factor analysis in the absence of cardio-metabolic risk.

		Correlation
F1	F2	-.342
F2	F3	.121
F3	F1	.166

Table 69 Confirmatory factor analysis variances in the absence of cardio-metabolic risk.

	Estimate	S.E.	C.R.	P	Label
F1	1.000				
F2	1.000				
F3	1.000				
e11	124.970	21.310	5.864	***	e1
e21	55.986	3.964	14.124	***	e2
e31	146.589	23.420	6.259	***	e3
e23	92.879	93.558	.993	.321	e4
e13	55.657	3.497	15.914	***	e6
e24	55.657	3.497	15.914	***	e6

Table 70 Squared multiple correlations: Confirmatory factor analysis in the absence of cardio-metabolic risk.

	Squared Correlations
VAR00065	.693
VAR00063	.043
VAR00076	.712
VAR00064	.267
VAR00060	.054
VAR00059	.188

.

Table 71 Confirmatory factor analysis (factor score weights) in the absence of cardio-metabolic risk.

	VAR00065	VAR00063	VAR00076	VAR00064	VAR00060	VAR00059
F2	.002	.000	.042	.013	-.002	-.003
F1	.010	.001	-.012	-.004	.022	.030
F3	.061	.009	.002	.001	.002	.002

Results and Discussion

Table 66 shows the intercepts of the Confirmatory Factor Analysis for the setting where the analysis was performed on the data in the absence of cardio-metabolic syndrome. The analysis shows that among the independent variables, variable Var00076 was the most important with the highest intercept (123.205). This variable had a significant effect on the cardio-metabolic syndrome factors. It must be remembered that the *mean* of the response variable when all the independent variables are equal to "0" is "the intercept". These intercepts were seen to contribute significantly to regression estimation for the prediction of the presence of cardio-metabolic factors. They all highly influence the three extracted factors. The significance of their contribution is quite high with p-values each of less than 0.001 for all the intercepts Hellsten (2006).

Table 69 indicates that systolic blood Pressure had a higher influence on factor 1 (F1) than the other variable, diastolic blood pressure. Systolic Blood pressure loaded with 5.377 on the first factor. This was highly significant with a p-value of 0.005, which is significant at the 2.5% and even 5% levels of significance. Diastolic blood Pressure was also significant at the 5% level of significance with a factor loading of 1.377 Byrne (2001), Hoffmann (2008).

The analysis also shows that Post Load Glucose had a higher influence on factor 2 than its counterpart, Glucose Fasting. Post Load Glucose heavily loaded on the F2 with a loading of 15.148. This heavy loading implied that compared to glucose fasting, post load glucose is more influential than glucose fasting. The high significance was with a p-value of less than 0.001 and only indicated by ***. Further, glucose fasting which loaded with 7.312 was also influential with a p-vale of less than 0.001.

Finally, Var00065 (waist circumference) with a heavy loading of 11.202 had the highest influence on factor 3 (F3). This influential variable on F3 was significant with a p-value of less 0.001 and only indicated in the output by ***. The heavy loading of waist circumference shows that waist circumference had a strong impact on the emergency of cardio-metabolic syndrome factors especially obesity. Obesity is one of the factors under cardio0metabolic conditions. The last variable

is BMI (Body Mass Index) which loaded with 1.584 with a p-value of 0.003. This was significant at the 0.025 and 0,05 levels of significance. The heavy loadings of the influential variables demonstrated the fact that they positively contribute to the manifestation of cardio-metabolic factors which are deadly disease conditions. The next step was to establish the degree of covariance between pairs of factors. Table 72 and 73 give the following covariance, correlation and the p-values between pairs of the extracted factors as indicated:

F1--F2 (.015), F2--F3 (.148), F3--F1 (.282). It was observed that for any change in F1, there will be a corresponding change in F2 and similarly a change in F2 will lead to a change in the same direction in F1. The corresponding p-value is 0.015 which is less than 0.05, a default level of significance. The existence of F1 was likely to lead to the existence of F2. The other two pairs did not seem to obviously predict the other. The corresponding correlations of the factors were stated here as: F1--F2 (-.342), F2--F3 (.121) and F3--F1 (.166). Similar the covariance values, the correlations demonstrate the same findings as those found under the covariance values.

The variances and their p-values (table 69 and 71) are indicated as follows: e11 (124.970, ***), e21 (55.986, ***), e31 (146.589, ***), e23 (92.879, .321), e13 (55.657, ***), e24 (55.657, ***). The large variances namely for the variables; e31, e11, e23 and e21 contribute to high level variability within the identified variable which may create some degree of confusion in the understating of variables.

Finally, it is estimated that the predictors of VAR00065 explain 69.3 percent of its variance. In other words, the error variance of VAR00065 is approximately 30.7 percent of the variance of VAR00065 itself. Also, it is estimated that the predictors of VAR00063 explain 4.3 percent of its variance. In other words, the error variance of VAR00063 is approximately 95.7 percent of the variance of VAR00063 itself. And that the predictors of VAR00076 explain 71.2 percent of its variance. In other words, the error variance of VAR00076 is approximately 28.8 percent of the variance of VAR00076 itself.

5.3.4. Factor Analysis for data Split According to the Presence of Metabolic Syndrome

This section presents Confirmatory Factor Analysis output in the presence of Metabolic Syndrome together with the required interpretation.

Table 72 Confirmatory factor analysis regression weights in the presence of cardio-metabolic risk.

		Estimate	S.E.	C.R.	P	Label
VAR00059	**F1**	15.6768	3.9652	3.9536	***	b1

VAR00060	**F1**	17.4851	4.2064	4.1568	***	b2
VAR00064	**F2**	127.7964	13.6026	9.3950	***	b3
VAR00076	**F2**	479.1575	47.3124	10.1275	***	b4
VAR00063	**F3**	103.0821	3.7203	27.7080	***	b6
VAR00065	**F3**	1.4344	1.0992	1.3049	.1919	b7

Table 73 Standardized regression weights: Confirmatory factor analysis (Group 1-Default model) in the presence of cardio-metabolic risk.

		Estimate
VAR00059	**F1**	.5557
VAR00060	**F1**	1.1112
VAR00064	**F2**	.8467
VAR00076	**F2**	1.1603
VAR00063	**F3**	.9894
VAR00065	**F3**	.0943

Table 74 intercepts by the default model in the presence of cardio-metabolic risk.

	Estimate	**S.E.**	**C.R.**	**P**	**Label**
VAR00059	127.9817	1.4142	90.5005	***	
VAR00060	79.3379	.7886	100.6012	***	
VAR00064	166.0410	7.5386	22.0255	***	
VAR00076	376.9503	20.8836	18.0501	***	
VAR00063	29.3334	5.2030	5.6378	***	
VAR00065	80.2989	1.0887	73.7540	***	

Table 75 Covariance's between extracted factors in the presence of cardio-metabolic risk.

		Estimate	**S.E.**	**C.R.**	**P**	**Label**
F1	**F2**	-.0919	.0520	-1.7666	.0773	
F3	**F2**	-.0012	.0320	-.0363	.9711	
F3	**F1**	.0302	.0458	.6589	.5100	

Table 76 Correlations in the presence of cardio-metabolic risk.

		Estimate
F1	**F2**	-.0919
F3	**F2**	-.0012
F3	**F1**	.0302

Table 77 Variance estimates in the presence of cardio-metabolic risk.

	Estimate	**S.E.**	**C.R.**	**P**	**Label**
F1	1.000				
F2	1.000				
F3	1.000				
e11	5502196	123.7548	4.4460	***	e1
e21	-58.1473	146.1635	-.3978	.6908	e2
e31	6451.1427	3148.6952	2.0488	.0405	e3
e23	-59069.1393	43963.7199	-1.3436	.1791	e4
e13	229.5486	23.2763	9.8619	***	e6
e24	229.5486	23.2763	9.8619	***	e6

Table 78 Squared multiple correlations: Confirmatory factor analysis in the default model in the presence of cardio-metabolic risk.

	Estimate
VAR00065	.0089
VAR00063	.9789
VAR00076	13464
VAR00064	.7168
VAR00060	1.2349

Table 79 Confirmatory Factor analysis (factor score weights) in the presence of cardio-metabolic risk.

	VAR00059	VAR00060	VAR00064	VAR00076	VAR00063	VAR00065
F1	.0000	-.0001	.0236	-.0575	.0213	-.0020
F2	.0000	-.0001	.0006	-.0014	.0796	-.0075
F3	.0001	.0095	.0000	.000	.0001	.0000

Results and Discussion

Table 72 indicates that Var00060 (diastolic blood pressure) with a loading of 17.4851 has a higher influence on factor 1 (F1) than Var00059 (systolic blood pressure) with a loading of 15.6768. The respective p-values for the two highly significant variables which loaded on the first factor are both less than 0.001. The output indicated the p-values only as ***. The significant loading on the first factor by the two variables meant that the two variables contributed to the manifestation of the condition of the factor of blood pressure (F1) figure 28 and figure 29. It is also noted that the Var00076 with a loading of 479.1575 had a higher influence on the second factor (F2) than its counterpart Var00064 with a loading of 127.7964. The two have less than 0.001 p-value each which makes them to be individually highly significantly influential to the emergency and existence of the second factor (F2).

Finally, Var00063 (body mass index) with a loading of 103.0821 had a higher influence on factor 3 (F3). However, the analysis showed that Var00065 (waist circumference) had no influence on factor three. The loading for waist circumference on factor 3 (F3) was 1.4344. This analysis meant that the body mass index was the most serious indicator of obesity (factor 3). Body Mass Index had a p-value of less than 0.001 Figure 28 and figure 29).

The respective intercepts and the corresponding p-values for different variables were indicated against the variables as follows: VAR00059 (127.9817, ***), VAR00060 (79.3379, ***), VAR00064 (166.0410, ***), VAR00076 (376.9503, ***), VAR00063 (29.3334, ***), VAR00065 (80.2989, ***). It is noted that some variables had higher intercepts than others, however, all the variables highly influenced cardio-metabolic factors. The high significance of the variable intercepts showed

that all the variables matter as far as the emergence and manifestation of cardio-metabolic factors are concerned Hellsten (2006).

The following factor combinations with their respective covariances and correlations were recorded for this analysis. The factors to factor covariance values; F1--F2 (-.0919, .0773), F3--F2 (-.0012, .9711), F3--F1 (.0302, .5100), and the correlation values F1--F2 (-.0919), F3--F2 (-.0012), F3-- F1 (.0302) are recorded. It is observed that the covariances were not significant with large p-values, which implied that the variables had little or no relationship as far as change of a variable relative to the change in another variable was concerned. Lack of significance in this case showed that there were no corresponding changes in the corresponding variables

Squared Multiple Correlations

It is estimated that the predictors of VAR00065 explain .89 percent of its variance. In other words, the error variance of VAR00065 is approximately 99.21 percent of the variance of VAR00065 itself while the predictors of VAR00063 explain 97.89 percent of its variance. In other words, the error variance of VAR00063 is approximately 2.11 percent of the variance of VAR00063 itself. Further, it is estimated that the predictors of VAR00064 explain 71.68 percent of its variance. In other words, the error variance of VAR00064 is approximately 28.32 percent of the variance of VAR00064 itself.

Referring to **Error! Reference source not found.**, two manifest variables (Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP)) were of research interest here. The two led to the construct F1. The output showed that when the F1 factor went up by 1 unit, Systolic Blood Pressure went up by 15.6768 units while by the same increase in F1, Diastolic Blood Pressure went up by 17.4851 units. The corresponding standard errors were 3.9652 and 4.2064 respectively. Division of estimates by the corresponding standard error yielded critical ratios given by 3.9536 for SBP and 4.1568 for Diastolic Blood Pressure. The probability of getting a critical ratio as large as 3.9536 in absolute value is less than 0.001. In other words, the regression weight for F1 in the prediction of Systolic Blood Pressure is significantly different from zero at the 0.001 level (two-tailed). It was also observed from the analysis that, the probability of getting a critical ratio as large as 4.1568 as an absolute value is less than 0.001.

Finally, the regression weight for F1 in the prediction of Diastolic Blood Pressure is significantly different from zero at the 0.001 level (two-tailed). Referring to table 80, the variables (Fasting

Glucose and Post Load Glucose) were related to factor F1 by the following interpretation: When F2 went up by 1 unit, Fasting Glucose went up by 127.7964 units while Post Load Glucose went up by 479.1575 units. The two had 13.6026 and 47.3124 as their respective standard errors. The respective critical ratios were 9.3950 and 10.1275 respectively for Fasting Glucose and Post Load Glucose. The flagging of *** at the end of each observe endogenous variable reflected highly significant p-values at the 0.001 level of significance. This means that the regression weight for F2 in the prediction of VAR00064 is significantly different from zero at the 0.001 level (two-tailed) and the regression weight for F2 in the prediction of VAR00076 is significantly different from zero at the 0.001 level (two-tailed).

The analysis and interpretation stated above confirm what was obtained under the Exploratory Factor Analysis where, overall, three factors were created as constructs. The researcher tried to confirm what was obtained under the Exploratory Factor Analysis. The results so far obtained were quite positive and thus encouraging. Thus in the presence of Metabolic Syndrome, the researcher has confirmed the results obtained under Exploratory Factor Analysis.

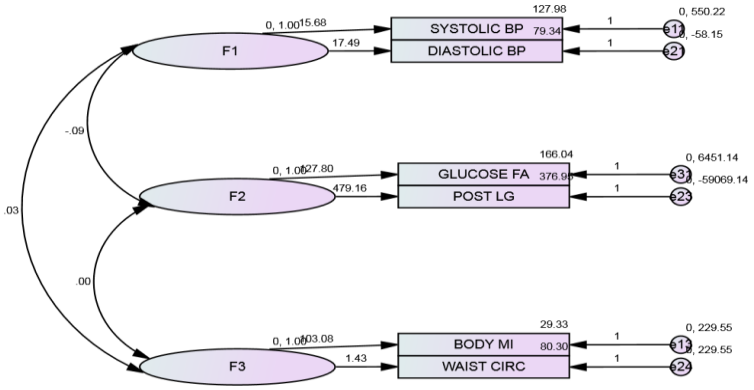

Figure 28 Unstandardized Factor Analysis according to the presence of Metabolic Syndrome in the general population.

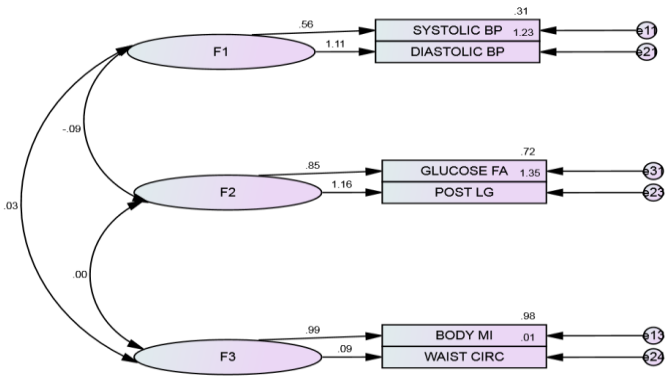

Figure 29 Standardized Factor Analysis in the presence of Metabolic Syndrome.

5.3.5. Confirmatory Factor Analysis of Data Split and Output for Females

This section presents Confirmatory Factor Analysis output for females together with the required interpretation.

Table 80 Regression weights among females analysis.

		Estimate	S.E.	C.R.	P	Label
VAR00059	**F1**	10.4251	.9595	10.8651	***	b1
VAR00060	**F1**	18.0647	.5352	33.7559	***	b2
VAR00064	**F2**	103.3126	3.6152	28.5770	***	b3
VAR00076	**F2**	226.0706	6.5803	34.3558	***	b4
VAR00063	**F3**	.3704	.0223	16.6383	***	b6
VAR00065	**F3**	-10.9404	.4879	-22.4214	***	b7

Table 81 Standardized regression weights for females in the default model.

		Estimate
VAR00059	**F1**	.4719
VAR00060	**F1**	1.3389
VAR00064	**F2**	.9399
VAR00076	**F2**	1.0368
VAR00063	**F3**	.0575
VAR00065	**F3**	-8620

Table 82 Intercepts for females.

	Estimate	S.E.	C.R.	P	Label
VAR00059	119.1427	.9707	122.7348	***	
VAR00060	71.7434	.5928	121.0208	***	
VAR00064	117.1678	4.8298	24.32596	***	
VAR00076	200.5396	9.6290	20.08266	***	
VAR00063	24.5947	.2831	86.8741	***	
VAR00065	88.1486	.5694	154..8225	***	

Table 83 Covariances among females.

		Estimate	S.E.	C.R.	P	Label

F1	**F2**	.1218	.0289	4.2165	***	
F3	**F2**	.8581	.0188	45.7496	***	
F3	**F1**	-.3560	.0340	-10.4669	***	

Table 84 Correlations for females among factors.

		Estimate
F1	**F2**	.1218
F3	**F2**	.8581
F3	**F1**	-.3560

Table 85 Variances of confirmatory factor analysis among females.

	Estimate	S.E.	C.R.	P	Label
F1	1.0000				
F2	1.0000				
F3	1.0000				
e11	379.4426	24.33327	15.5939	***	
e21	-144.2903	20.1657	-7.1552	***	
e31	1408.5398	98.2271	14.3396	***	
e23	-3563.9448	2+62.6783	-13.5677	***	
e13	41.3807	2.5717	16.0910	***	
e24	41.3807	2.5717	16.0910	***	

Table 86 Squared multiple correlation: Factor analysis among female respondents.

	Estimate
VAR00065	.7431
VAR00063	.0033
VAR00076	1.0750
VAR00064	.8834
VAR00060	1.7926
VAR00059	.2227

Results and Discussion

Table 80 indicates that variable Var00060 (DBP) which had a loading of 18.0647 on factor 1 (F1), and a p-value less than 0.001, had a high influence on factor 1 (Blood Pressure). Also, variable Var00059 (SBP) with a loading of 10.4251 had a high influence on factor 1. The two variables which had significant individual and independent influences on factor 1 implied positive contributions by the two variables both combined and independently. Systolic blood pressure and diastolic blood pressure had direct and indirect independent influences on the elevation of blood pressure.

Also, still from Table 80, VAR00076 (Post load glucose) with a loading of 226.0706 had a high influence on the second factor (F2) (Glucose). Post load glucose had a p-value of less than 0.001. The other variable which had a high influence on factor 2 was Var00064 with a loading of

103.3126. This variable also had a negligible p-value less than 0.001 signifying its high significance. The two variables being highly significant showed that the promotion of the two variables promoted diabetes in individuals. Elevated blood sugar has been known to cause diabetes in humans, a member of the family of cardio-metabolic factors.

Thirdly, Table 80 showed that Var00065 with a loading of -10.9404 and a p-value of less than 0.001 had a high influence on factor 3 (F3) (obesity). The other variable with a loading of 0.3704 with a p-value of less than 0.001 had a similar high loading on factor 3. The small p-values for the two variables indicate that they had direct and indirect effects on the elevation of obesity, one of the factors under cardio-metabolic syndrome constellation.

Table 82 shows the intercepts for the analysis of these data. The intercepts are listed here against the variables and the corresponding p-values as follows: VAR00059 (119.1427. ***), VAR00060 (71.7434, ***), VAR00064 (117.1678, ***), VAR00076 (200.5396, ***), VAR00063 (24.5947, ***), VAR00065 (88.1486, ***). The intercepts so recorded show very high values which demonstrate how significantly they contribute to the cardio-metabolic syndrome factors. With regard to variability, the following were the recorded covariances and correlations between selected pairs of the factors extracted from the data set (table 88 & 89):

Tables 83 and 84 give the covariances and correlations including indications of p-values, F1--F2 (.1218, ***), F3--F2 (.8581, ***), F3--F1 (-.3560, ***) and observed correlations F1--F2 (.1218), F3--F2 (.8581), and F3--F1 (-.3560). From the analysis, it could be seen that the variables were neither related nor did they show significant p-values. The failure by the variables to show any significant covariances demonstrate lack of promotion of cardio-metabolic factors within the female gender of participants.

Table 85 gives variances and their respective p-values for the indicated variables within the model: e11 (379.4426, ***), e21 (-144.2903, ***), e31 (1408.5398, ***), e23 (3563.9448, ***), e13 (41.3807, ***), e24 (41.3807, ***). It was noted that the variances were so large in magnitude as well as being significant that the variables had varying data in such a manner that the proper distribution of the data was unclear.

Discussion of Results:

Referring to Table 80, this analysis and interpretation was performed when the data were split on the basis of gender (whether male or female). This specific section dealt with the case where the data were obtained on female participants. The basic assumption was that the above manifest observed variables were directly related to the determined factor F1. According to the Exploratory Factor Analysis, the researcher was able to identify F1 as the Blood Pressure construct. The Confirmatory Factor Analyses leads to the following interpretation: Whereas when Blood Pressure (F1 factor) went up by 1 unit, Systolic Blood Pressure went up by 10.4251 units while Diastolic Blood Pressure went up by 18.0647. Respectively, 0.9595 and 0.5352 were the observed standard errors for SBP and DBP leading to critical values determined to be 10.8651 and 33.7559. The corresponding p-values were indicated by *** for both the tests. This meant that the probability of getting critical ratios as large as 10.86511 for SBP and 33.7559 for DBP in absolute values was less than 0.001. That is to say, the regression weights for F1 in the prediction of Systolic Blood Pressure and Diastolic Blood Pressure were significantly different from zero at the 0.001 level (two-tailed).

5.3.6. Confirmatory Factor Analysis of Data Split for Males

This section presents Confirmatory Factor Analysis output for males together with the required interpretation.

Table 87 Regression weights among males.

		Estimate	S.E.	C.R.	P	Label
VAR00059	**BP**	21.4386	4.6389	4.6214	***	b1
VAR00060	**BP**	6.9464	1.5910	4.3659	***	b2
VAR00064	**Glucose**	87.6653	7.4544	11.7603	***	b3
VAR00076	**Glucose**	218.5658	16.9501	12.8946	***	b4
VAR00063	**Obesity**	96.7173	3.2693	29.5823	***	b6
VAR00065	**Obesity**	1.4935	.6762	2.2088	.0272	b7

Table 88 Regression weights under confirmatory factor analysis for males.

		Estimate
VAR00059	**BP**	1.0077
VAR00060	**BP**	.5311
VAR00064	**Glucose**	.8561
VAR00076	**Glucose**	1.1214.
VAR00063	**Obesity**	.9893
VAR00065	**obesity**	.1041

Table 89 Standardized regression weights for males using confirmatory factor analysis.

	Estimate	S.E.	C.R.	P	Label
VAR00059	119.3731	.9983	119.5711	***	
VAR00060	71.4394	.6139	116.3793	***	
VAR00064	117.1753	4.7907	24.4591	***	
VAR00076	197.2450	9.1959	24.4493	***	
VAR00063	28.6517	4.5730	6.2654	***	
VAR00065	80.0371	.6708	119.3170	***	

Table 90 Covariance's of confirmatory factor analysis for males.

		Estimate	S.E.	C.R.	P	Label
BP	**Glucose**	.1279	.0539	2.3711	.0177	
Obesity	**Glucose**	.0238	.0370	.6425	.5206	
Obesity	**BP**	-.0423	.0479	-.8845	.3764	

Table 91 Correlations of confirmatory factor analysis over factors among males.

		Estimate
BP	**Glucose**	0.1279
Obesity	**glucose**	.0238
obesity	**BP**	-.0423

Table 92 Variances of confirmatory factor analysis for males.

	Estimate	S.E.	C.R.	P	Label
BP	1.0000				
Glucose	1.0000				
Obesity	1.0000				
e11	-7.0400	196.6246	-.0358	.9714	
e21	122.8301	22.1940	5.5344	***	
e31	2799.5714	1138.2497	2.4595	.0139	
e23	-9780.5461	7005.9995	-1.3960	.1627	
e13	203.4034	13.4560	15.1162	***	
e24	203.4034	13.4560	15.1162	***	

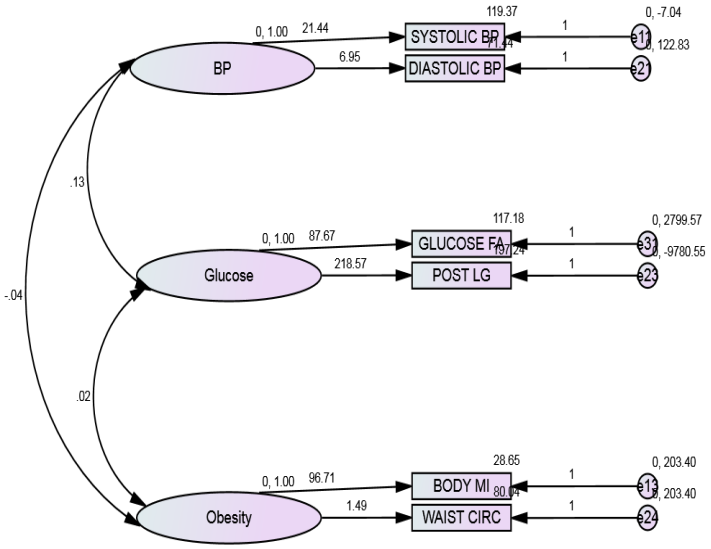

Figure 30 Unstandardized Factor Analysis Estimates among males.

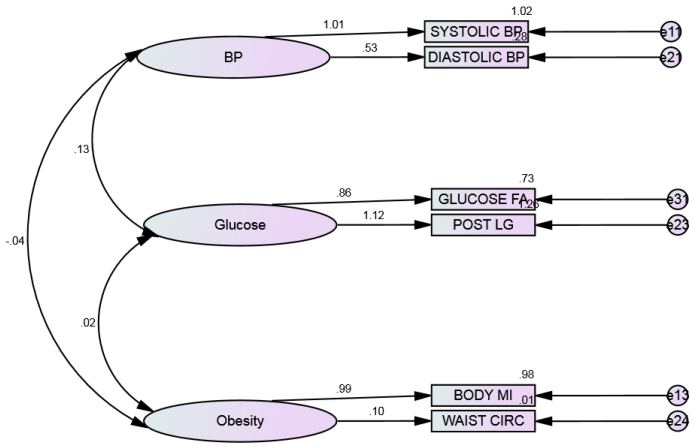

Figure 31 Standardized Factor Analysis Estimates among males.

Figures 30 and figure 31 indicate that among males for this analysis, Var00059 with a loading of 21.4386 and a p-value of less than 0.001 had a higher influence on factor 1 (BP) than variable Var00060. Var00060 had a loading of 6.9464 with a p-value of less than 0.001. The indications of

the two p-values were only shown as ***, signifying high level significance with a p-value less than 0.001. The analysis also shows that Var00060 had a high influence on factor 1 (BP). The two variables having highly significant influences either both combined or separately and independently on blood pressure shows that the two variables have direct positive effects on blood pressure, one of the factors of the cardio-metabolic diseases. It remains to see the relationship between the two factors later in the interpretation using covariances and correlations.

In addition to the above analysis and interpretation, variable Var00076 with a loading of 218.5658 and a p-value of less than 0.001 had a higher influence on factor 2 than variable Var00064 which had a loading of 87.6653 with a p-value of less than 0.025. The two variables, however had strong impacts on their influences on factor 2 (Glucose). This analysis brings to light the fact that the two variables being post load glucose and glucose fasting, have strong bearings on the elevation of the diabetes in humans. Either a combined or independent effects by the two variables could be disastrous to health as the increase in levels of either could lead to the manifestation of type 2 diabetes mellitus, a cardio-metabolic condition of destructive nature. Thus control of the levels of both is important.

Thirdly, figures 30 and 31 and tables 87 and 88 indicate that Var00063 with a loading of 96.7173 and a p-value of less than 0.001 had a higher influence on factor 3 (obesity) than variable Var00065 with a loading of 1.4935 and a p-value of 0.0272. This analysis showed that variable Var00065 had also a high and significant influence on obesity. Obesity is one of the causes of factors of metabolic syndrome. The high loading on this factor and the highly significant p-values show that high levels of the two variables can increase the obesity factor to dangerous levels. The two variables forming this factor were body mass index and waist circumference. Thus, the increase of body mass index has been established to lead to serious health risks.

Tables 95 and 96 show covariance values and correlations for the factors extracted from the data under this investigation. The following list populates the observed variables and the corresponding covariance value and the p-value. At the end of the covariance estimates, bivariate correlations have been provided. BP—Glucose (.1279, .0177), Obesity—Glucose (.0238, .5206), Obesity—BP (-.0423, .3764), BP—Glucose (0.1279), Obesity—glucose (.0238) and obesity—BP (-.0423).

First, it is observed that due to obvious formula reasons, there exists a very strong relationship between the covariance values and the observed correlations. This means that any conclusion

drawn on a covariance value will equally apply to a correlation value. Some covariances were observed to be negligible. For covariances with significant p-values, we note that the involved factors have a positive relationship which indicates that the two variables either increase together or decrease together at a significant rate. Considering the output once again, BP and Glucose were found to be positively and significantly related with an observed p-value of 0.0177. The overall conclusion was that BP and Glucose either combined or independently can manifest as a cardio-vascular condition. This means that blood pressure and diabetes were likely to emerge at the same time or possibly at different times. According to the results, obesity and BP have a low but negative relationship meaning that they may not necessarily occur at the same time.

Finally, the following table gives the variances and the corresponding p-values for latent variables in the model (table 97): e11 (-7.0400, .9714), e21 (122.8301, ***), e31 (2799.5714, .0139), e23 (-9780.5461, .1627), e13 (203.4034, ***), e24 (203.4034, ***). The majority of the variances have confirmed to be insignificant whereas a few others were significant.

5.4. Comparison of Communalities under different Extraction Methods

The variables under this comparison were; Glucose Fasting, Post Load Glucose, Systolic Blood Pressure, Diastolic Blood Pressure, Waist Circumference and Body Mass Index. The thesis research data consisted of many variables but the researcher chose to utilise these selected variables in order to demonstrating the current research idea under this topic.

It is an observation that generally, different extraction procedures produced different communalities. Seven extraction methods were compared under the same (Promax) rotation procedure. Four of the rotation extraction methods (Principal Component Analysis, Principal Axis Factoring, Alpha Factoring, and Image Factoring) produced smaller communalities while the rest three (Un-weighted Least Squares, Generalised Least Squares, Maximum Likelihood) produced higher communalities. A further analysis reveals that even those extraction methods, which had higher communalities, were not consistent across all the variables. Some variables had very low communalities while others had very high communalities under the same extraction method. This leaves a lot to be desired from the research point of view. The "big" question is "what influences the magnitude of a communality for a given variable?" The only answer available for now is that there is need for research to establish this occurrence.

It can be concluded that there can be no rule of thumb of the choice of an extraction method. The extraction method will depend on the variable being analysed.

Table 93: A Comparison of Different Extraction Methods

	Extraction Method						
Variable	**Principle Component**				**Un-weighted Least Squares**		
	Factor 1	Factor 2	Factor 3		Factor 1	Factor 2	Factor 3
POST LG	.871				.999		
GLUCOSE FA	.862				.507		
DIASTOLIC BP		.858				.998	
SYSTOLIC BP		.858				.479	
WAIST CIRC			.764				.999
BODY MI			.727				.113
	Generalised Least Squares				**Maximum Likelihood**		
	Factor 1	Factor 2	Factor 3		Factor 1	Factor 2	Factor 3
POST LG	.951				.960		
GLUCOSE FA	.505				.500		
DIASTOLIC BP		.888				.934	
SYSTOLIC BP		.573				.544	
WAIST CIRC			1.000				1.000
BODY MI			.116				.116
	Principal Axis Factoring				**Alpha Factoring**		
	Factor 1	Factor 2	Factor 3		Factor 1	Factor 2	Factor 3
POST LG	.736				.764		
GLUCOSE FA	.690				.666		
DIASTOLIC BP		.715				.720	
SYSTOLIC BP		.676				.664	
WAIST CIRC			.550				.479
BODY MI			.200				.246
	Image Factoring						
	Factor 1	Factor 2	Factor 3				
POST LG	.450						
GLUCOSE FA	.446						
DIASTOLIC BP		.416					
SYSTOLIC BP		.414					
WAIST CIRC			.111				
BODY MI			.102				

A comparison of extraction methods under the same approach shows that the loadings were a little tricky. Analysed according to factors, one observes the first factor (among all the extraction methods) had the highest loading as compared to loadings of other factors. Furthermore, it can be noticed that for a given factor, the highest loading was for the first variable. This is justifiable

following the analytical procedures of extraction of factors, which give the first factor the highest loading. Using the above observation, one notices that the "Un-weighted Least Squares" had the highest loadings on the first variable under all the factors. However, Principal Component Analysis had the best overall extraction procedure where all the variables under all the extracted factors had high loadings. "Image Factoring" had the lowest loadings across the factors and even across the variables. This means that selecting the Un-weighted Least Squares will be a big risk due to unpredictable nature of the loading structure. The best bet would be "Principal Component Analysis" which is comparatively more reliable due to its consistency.

5.5. Association of FA to other Multivariate Methods

A further analysis was performed on the data to compare FA to other multivariate methods and analyse their association. Discriminant analysis was selected amongst many to compare its association to FA. The analysis carried out was as follows;

The objective of the discriminant analysis for these data

All the required conditions for performing discriminant analysis were satisfied except the one where the p-value was significant.

This type of analysis allows one to determine the probability of group membership based on predictor variables. Under this analysis, the independent variables are predictors while the dependent variable consists of groups.

Assumptions

Independent predictor variables, group membership must be mutually exclusive. There must be no outliers. Predictors should be normally distributed; within group variance in the covariance matrix should be equal across groups. Predictor variables should not be highly correlated with one another (multi-collinearity).

Tests of Equality of Group Means

The figure below shows the test of equality of group means. It can be observed that those means, each with a p-value less than 0.05 were good predictors of Metabolic Syndrome. It means further, that under this analysis, Systolic Blood pressure, Diastolic Blood pressure, Glucose Fasting and Post

Load Glucose are good predictors of metabolic syndrome. These variables have p-values far less than 0.05. Other variables have p-values greater than 0.05 and thus they are poor contributors to Metabolic Syndrome.

Table 94: Test of Equality of Group Means – Pooled within group matrices

	Wilks' Lambda	**F**	**df1**	**df2**	**Sig.**
Systolic BP	.883	49.321	1	372	.000
Diastolic BP	.753	121.973	1	372	.000
Body	.994	2.299	1	372	.130
Glucose FA	.904	39.705	1	372	.000
Waist Circ	1.000	.177	1	372	.674
Post LG	.805	90.383	1	372	.000

Correlations are not too large. Bivariate correlations are observed to be less than or equal to 0.40. The smallest correlation was 0.000, which was between diastolic blood pressure whereas the highest correlation (0.374) was between diastolic blood pressure and systolic blood pressure. This is quite expected due to the nature of the two blood pressures, where, as one goes up, there is always a tendency of corresponding increase of the other.

Table 95: Box's Test of Equality of Covariance Matrices

		SYSTOLIC BP	**DIASTOLIC BP**	**BODY MI**	**GLUCOSE FA**
Correlation	**SYSTOLIC BP**	1.000	.374	-.079	-.166
N	**DIASTOLIC BP**	.374	1.000	.029	-.201
	BODY MI	-.079	.029	1.000	.060
	GLUCOSE FA	-.166	-.201	.060	1.000
	WAIST CIRC	.002	.000	.120	.071
	POST LG	-.125	-.273	.074	.435

Log Determinants

The idea of Log Determinant is to ensure that group means are more-or-less equal. According to the analysis, the Log Determinant shows that group means are approximately equal. This means that the Test of Equality of Covariance Matrices support the hypothesis that the observed Log Determinant are approximately equal. This is evidenced by the closeness of the Log Determinant values are more or less equal.

Table 96: Log Determinants.

CARDIOMETABOLIC RISK	Rank	Log Determinant
Presence	6	42.157
Absence	6	29.137
Pooled within-groups	6	37.888

Box's M

The observed p-value should be higher than 0.05. The idea is to fail to reject the null hypothesis of equal population covariance matrices. The observed p-value was 0.000 far less than the lowest default one of 0.05. The outcome shows that the test's null hypothesis of equal population covariance matrices should be rejected implying that the population covariance matrices are equal.

Table 97: Test Results

Box's M		1875.131
F	Approx.	87.293
	df1	21
	df2	159917.185
	Sig.	.000

Summary of Canonical Discriminant Functions

Effects size

The effect size for this analysis was observed to be the (square of 0.681) which gave 0.464, meaning that the effect size was 46.4%. First 1 canonical discriminant functions were used in the analysis.

Table 98: Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.865a	100.0	100.0	.681

Wilks' Lambda

The observed Wilks' Lambda (0.536) is significant with a p-value (0.000), less than 0.05, which shows that the model is good. The observed p-value being significant shows that the prediction model is statistically significant.

Table 99: Wilks' Lambda.

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.536	230.045	6	.000

Standardized Canonical & Structure matrices

The two matrices must show consistency. Diastolic BP (0.757 for discriminant function coefficients) and Diastolic BP (0.616 for the structure matrix). Also, Post LG (.652 for discriminant function coefficients) and Post LG (.530 for the structure matrix). Other values namely Systolic BP (.235 for discriminant function coefficients), Systolic BP (.391 for the structure matrix), Glucose FA (.262 for discriminant function coefficients) and Glucose FA (.351 for discriminant function coefficients). Accordingly, those variables with higher loadings are better predictors of the independent variable, metabolic syndrome. This outcome is interpreted to mean that the best predictor of the dependent variable is Diastolic Blood Pressure, followed by Post LG. This analysis shows that the two matrices are consistent. The worst predictor is observed to be Waist Circumference, which had -0.070 for discriminant function coefficients while Waist Circumference had -0.023 for the structure matrix.

Table 100: Standardized Canonical Discriminant Function Coefficients

	Function
	1
SYSTOLIC BP	.235
DIASTOLIC BP	.757
BODY MI	.026
GLUCOSE FA	.262
WAIST CIRC	-.070
POST LG	.652

Table 101: Structure Matrix

	Function
	1
DIASTOLIC BP	.616
POST LG	.530
SYSTOLIC BP	.391
GLUCOSE FA	.351
BODY MI	.085
WAIST CIRC	-.023

Table 102: Standardized Canonical Discriminant Function Coefficients.

	Function
	1
SYSTOLIC BP	.013
DIASTOLIC BP	.071
BODY MI	.000
GLUCOSE FA	.017
WAIST CIRC	-.005
POST LG	.014

Classification Statistics

Table 103: Prior Probabilities for Groups

CARDIOMETABOLIC RISK	Prior	Cases Used in Analysis	
		Unweighted	Weighted
Presence	.286	107	107.000
Absence	.714	267	267.000
Total	1.000	374	374.000

In the Presence metabolic syndrome, the following coefficients were recorded: .329, .721, -.006, .414, .358, .073 and the intercept -90.814. If we regard Y be to be the response and X_i be the ith predictor, then the Response for the presence of metabolic syndrome will be denoted by: If Y_p can be regarded to be the response and Y_i be the i^{th} coefficient, then the response will be denoted by:

$$Y_p = -90.814 + 0.329X_1 + 0.721X_2 - 0.006X_3 + 0.414X + 0.358X_4 + 0.03X_5,$$

Which serves to represent the model for the presence of metabolic syndrome. The presence of metabolic syndrome depends on those predictor variables with high coefficients. The model for the absence of metabolic syndrome is equally represented by:

$$Y_A = -69.515 + 0.302X_1 + 0.575 - 0.006X_3 + 0.380X + 0.368X_4 + 0.045X_5,$$

Which serves to represent the model for the absence of metabolic syndrome.

Table 104: Classification Function Coefficients by using Fishers' Linear Discriminant Function

	CARDIOMETABOLIC RISK	
	Presence	Absence
SYSTOLIC BP	.329	.302
DIASTOLIC BP	.721	.575
BODY MI	-.006	-.006
GLUCOSE FA	.414	.380
WAIST CIRC	.358	.368
POST LG	.073	.045
(Constant)	-90.814	-69.515

Classification Results

The figure below shows the percentage of group membership which define the degree of sensitivity and specificity. The analysis shows that the sensitivity was observed to be 70.1% while the specificity was 98.3%. This analysis shows that among those who had the cardiometabolic syndrome, 70.1% were correctly predicted to have metabolic syndrome.

Overall, the analysis demonstrates that 86.7% of original grouped cases correctly classified whereas 83.7% of cross-validated grouped cases correctly classified. These results show very high percentages of sensitivity and specificity. These results talk a positive story regarding the data.

Table 105: Classification Results

CARDIOMETABOLIC RISK			Predicted Group Membership		Total
			Presence	Absence	
Original	Count	Presence	282	120	402
		Absence	10	565	575
	%	Presence	70.1	29.9	100.0
		Absence	1.7	98.3	100.0
Cross-validated	Count	Presence	253	149	402
		Absence	10	565	575
	%	Presence	62.9	37.1	100.0
		Absence	1.7	98.3	100.0

Link between Factor Analysis and Discriminant Analyses

Factor Analysis and discriminant Analysis play complementary roles when applied to data of interest. For these data, factor analysis identified the number of factors and went further to identify the variables, which constituted the identified factors. For the data at hand, three factors were determined: Blood Pressure formed by systolic blood pressure and diastolic blood pressure, blood glucose metabolism disordering formed by Post Load Glucose and glucose fasting and obesity formed by Body Mass Index and waist circumference.

Under discriminant analysis, it was revealed that some of the variables were better contributors to the dependent variable (cardiometabolic syndrome) than others. At this point, discriminant showed that Diastolic BP (0.757 for discriminant function coefficients) and Diastolic BP (0.616 for the structure matrix). Also, Post LG (.652 for discriminant function coefficients) and Post LG (.530 for the structure matrix). Other values namely Systolic BP (.235 for discriminant function coefficients), Systolic BP (.391 for the structure matrix), Glucose FA (.262 for discriminant function coefficients) and Glucose FA (.351 for discriminant function coefficients). Accordingly, those variables with higher loadings are better predictors of the dependent variable (metabolic syndrome). This outcome is interpreted to mean that the best predictors of the dependent variable is Diastolic Blood Pressure, followed by Post LG. the degree of importance depending on the value of the coefficient as observed. These results have been supported by (Sinclair & Alexson, 1986).

In summary, whereas factor analysis is a precursor to data, discriminant on the other hand plays the role of establishing the degree of importance of the predictor variable(s). The two approaches paly complementary roles.

5.6. Conclusion

This thesis used factor analysis to define the clustering of non-lipid variables explaining CDM in Bantu Africans where a constructive development was successfully effected to identify factors such as; glucose metabolism, Blood pressure and Obesity. Since 3 factors in the sequence of dyslipidaemia, hypertension, and abdominal obesity-dysglycemia were identified for the Bantu Central Africans, MS phenotype, one more major factor could be accounted for by this specific MS. Early prevention and management (diagnosis and proper intervention) strategies for those modifiable loaded risk variables could reduce the burden of type 2 DM, MS, and emerging cardiovascular diseases in Central Africa.

This thesis was intended to determine component structures for seven selected variables from a dataset containing several variables received from Kinshasa Hinterland, in the Democratic Republic of Congo. The researcher chose to utilize seven variables which he believed could demonstrate the power of Factor Analysis in the analysis of medical data and by extension, other qualifying sets of data. The researcher applied both methods of Factor Analysis namely, Exploratory Factor Analysis and Confirmatory Factor Analysis. Each of the two procedures produced results which have been recorded and interpreted.

According to the EFA, results obtained showed different outcomes for different data treatments. In the whole population, and in the presence of Cardio-metabolic Risk, this research revealed an eigenvalue oriented factor formation. There, the emphasis of fighting against Metabolic Syndrome drew more attention on Blood Glucose Metabolism Disordering than on other life factors determined. Blood Glucose Metabolism Disordering was followed in the order of seriousness by obesity. Analysis performed in males in the presence of Cardio-metabolic Risk revealed similar findings with almost the same loadings. This implied that the seriousness of Metabolic Syndrome as was observed by this study was on equal footing both in the presence of the general population as it was found to be in males in the presence of Cardio-metabolic Syndrome. Average anthropometric measurements dropped drastically among females in the absence of Metabolic Syndrome. Under this transformational setup, Confirmatory Factor Analysis produced interesting results. First, the

CFA results totally agreed with those established under EFA. The multivariate regression analysis showed heavy and highly significant regression weights between the exogenous and endogenous variables even when treated at the bivariate level.

All these summaries can be more clearly understood from the analysis and interpretation provided under every relevant section and subsection. Factor analysis performed for this study suggests that the clustering of the non-lipid variables is sufficient to define CDM in Bantu Africans where the research developed factors such as; Blood Glucose metabolism Disordering, Blood pressure and Obesity, (Nasila et al., 2013). Since 3 factors in the sequence of dyslipidaemia, hypertension, and abdominal obesity-dysglycemia were identified for the Bantu Central Africans, MS phenotype, one more major factor could be accounted for by this specific MS. Early prevention and management (diagnosis and proper intervention) strategies for those modifiable loaded risk variables could reduce the burden of type 2 DM, MS, and emerging cardiovascular diseases in Central Africa (Nasila at al., 2013).

Chapter 6

6. Conclusions and Recommendations

This chapter presents conclusions based on the analysis covered in Chapter 5 and also recommendations based on the conclusions in Chapter 6 .

6.1. Conclusions

Explanatory Factor Analysis

This study has brought out several interesting findings which were both statistical and medical oriented defined in all the study settings. Under the Exploratory Factor Analysis, Systolic Blood Pressure had a positive but significant relationship with Diastolic Blood Pressure. Post Load Glucose had a strong and positive relationship with Fasting Glucose (Chapter 5 table 2) showing that the positive effect by one of them on Cardio-metabolic Syndrome implies a similar effect by the other variable as evidenced by Nasila et al. (2013). Another pair of variables which had a significant correlation coefficient was Height and Body Mass Index. This analysis meant that in the general population, these variables had direct and either independent or combined effects on cardio-metabolic syndrome factor manifestation. Under this setup, the extracted factors explained close to 80% of the variance in BMI and 74% in Systolic Blood Pressure as shown by Nasila et al. (2013). The analysis showed a good model with the exception of Waist Circumference which had a high p-value. Three factors were extracted (Chapter 5 table 4) as evidenced by Wu et al. (2006). This was supported by the results of the scree plot (Chapter 5 figure 10) and Component plot in Rotated space (Chapter 5 figure 12) and table 5 which represented the loadings in a rotated component matrix and figure 13. In the absence of Cardio-Metabolic Risk, Systolic Blood Pressure was strongly correlated with Body Mass Index and Waist Circumference with significant positive correlations, Body Mass Index and had strong correlations with Systolic Blood Pressure and Diastolic Blood Pressure (Chapter 5 table 7 and figure 13).

In the absence of Cardio-Metabolic Syndrome, two factors were extracted and identified as; Glucose (Blood Glucose metabolism Disordering) and Obesity constituting of the variables, Body mass Index and Waist Circumference (Chapter 5 Table 7, figure 14, figure 15). In the presence of Cardio-metabolic Risk, all variables had significantly high means signaling a contribution towards

the existence and manifestation of Cardio-Metabolic Risk. Three factors were formed (Chapter 5 table 12, figure 16). Three factor extraction was supported by the Rotated Component Matrix (Chapter 5 table 14) and the Component Plot in Rotated Space (Chapter 5 figure 17). There, the correlations between the observe variables DBP and SBP, FG and PLG were very high (Chapter 5 table 13).

Under another scenario, where the data was composed of males in the presence of Cardio-metabolic Risk, the observed means were very high e.g., BMI (43.37) SBP (127.62) PLG (173.81). the correlations were noted to be significantly high between SBP & DBP, FG & PLG and WS & BMI. Further, the communalities were high showing the interrelatedness with the presence of Cardio-metabolic Risk factors. Under this setting, three identifiable factors; Blood Glucose Metabolism Disordering, Blood Pressure and Obesity with remarkably high readings. This showed a strong influence on the manifestation of Cardio-Metabolic factors (Chapter 5 table 21, table 22 and table 23 figure 18) as discussed by Shen et al. (2002).

In the absence of Cardio-Metabolic Syndrome, all the means for all the observed variables had a drastic drop. This sharp drop signified a positive health status of the participants. An SBP with a registered mean of 113.32, a PLG average reading of 122.11are examples of normal health anthropometric recordings. Under this setting once again, observed correlations between BMI & HT, FG & PLG were high but positive, low percentage explained by the extracted variables under communalities (table 28). Two factors were extracted (Chapter 5 Figure 19, table 29, table 30) (Chapter 5 figure 20) which showed the Component Plot in Rotated Space. Table 29 showed the component matrix as evidenced by Nasila et al. (2013).

Other situations were described by data split for females in the absence of Cardio-Metabolic Syndrome factors. Under this data setup, the observed variable means were low, the correlations between PLG & FG was too low, but between WC & HT, WC & DBP, WC & BMI, they were high but did not pose any meaningful risk due to poorly defined Cardio-metabolic factor risk combinations as discussed by Gurka et al. (2012) and Nasila et al. (2013). It meant that among females without metabolic risk factors, the participants exhibited normal health conditions. The communalities were all at 100%. Under this setup again, only two were well defined factors were extracted (Chapter 5 Table 35, figure 21). This finding was further supported by the component Plot in Rotated Space (Figure 22).

Other analyses such as males in the presence of Cardio-metabolic Syndrome and in rural setups demonstrated high correlations between pairs of defined variables. The high correlation under this setup signified high indications of the risks of Cardio-metabolic Syndrome factors and the possible presence of the risk factors. Here, 63.794% of the variance was explained by the extracted factors. Three factors were extracted (Chapter 5 table 39, figure 23). These results were supported by the component matrix for the data from a rural setup (Chapter 5 table 40), and the Component Plot in Rotated Space (Chapter 5 figure 24).

Confirmatory Factor Analysis

Under Confirmatory Factor Analysis, the analysis showed highly significant means and intercepts, showing the presence showing a positive influence on the occurrence of Cardio-metabolic Risk factors with high significant p-values (Chapter 5 table 56 & table 57). This was in the group-1, default model. The magnitude of the intercepts were so large that an indication of significant effects to the development of cardio-metabolic risk factors was evidenced. From the estimation point of view, the regression estimates were of mixed standards whereby, the following were observed: FG has a significant influence on SBP (table 54) and (Chapter 5 figure 26), BMI had no influence on WC, WC had no influence of DBP (Chapter 5 table 54 & figure 26), BMI had no influence on either DBP or SBP (Chapter 5 table 54 & figure 26). Furthermore, some variables had strong positive influences on Cardio-metabolic risk factors while others did not. FG, DBP had higher influences on the development of cardio-metabolic risk factors than others such as: SBP, WT, HT and WC (Chapter 5 table 54, figure 26).

When all diabetics was the exogenous variable in the model, GF had a higher influence on DBP than other variables, GF had also higher influence on diabetics than other variables. Also DBP had a high influence on all diabetics. All other observed variables has no influence on diabetics. Using CFA, for the unstandardized regression analysis, three factors, extracted under the EFA. The CFA analysis conformed that the extraction of the factors at the EFA level was quite genuine and thus confirmed that: SBP and DBP formed the first factor (F1), GF and PLG formed factor 2 (F2) and BMI and WC formed the third factor (F3). This confirmation proved the strength of factor analysis and hoe the two procedures can integrate and co-exist. The heavy loading of the two pairs of variables under every factor case signified the strong and high influence the two pairs of variables can have on the development of the respective factors in the model (Chapter 5 tables 54-57, figure 26).

In the presence of Cardio-metabolic Risk, the observed variables had highly influential means and intercepts to the development of Cardio-metabolic Risk factors owing to large value for both similar to Anderson et al. (2001). Other output showed covariances and correlations which were not significant Systolic (Chapter 5 tables 77-83, figure 28-29). Systolic Blood Pressure and Diastolic Blood Pressure heavily loaded on factor 1 (F1) identified as Blood Pressure, Fasting Glucose and Post Load Glucose heavily loaded on factor 2 (Blood Glucose Metabolism Disordering) and Body Mass Index (BMI) and Waist Circumference)WC) heavily loaded on the third factor namely, Obesity (Chapter 5 table 77, figure 28, 29).

Among females, all intercepts were highly influential towards the growth and development of cardio-metabolic syndrome owing to the large recorded values. Covariances between pairs of factors were similarly large showing the unpredictable nature of the manifestation of metabolic risk factors which may fluctuate between very large and very small observations (Chapter 5 tables 92-97, figures 30, 31).

Comparison of Different Rotation Methods

Following the detailed discussion in section 5.4. A comparison of extraction methods under the same approach shows that the loadings were a little tricky. Analysed according to factors, one observes the first factor (among all the extraction methods) had the highest loading as compared to loadings of other factors. Furthermore, it can be noticed that for a given factor, the highest loading was for the first variable. This is justifiable following the analytical procedures of extraction of factors, which give the first factor the highest loading. Using the above observation, one notices that the "Un-weighted Least Squares" had the highest loadings on the first variable under all the factors. However, Principal Component Analysis had the best overall extraction procedure where all the variables under all the extracted factors had high loadings. "Image Factoring" had the lowest loadings across the factors and even across the variables. This means that selecting the Un-weighted Least Squares will be a big risk due to unpredictable nature of the loading structure. The best bet would be "Principal Component Analysis" which is comparatively more reliable due to its consistency.

Factor Analysis and Other Methods

In summary, as discussed in detail in section 5.5. Whereas factor analysis is a precursor to data, discriminant on the other hand plays the role of establishing the degree of importance of the predictor variable(s). The two approaches paly complementary roles.

6.2. Recommendations

The recommendations outlined in this subsection of the chapter have been categorized according to the class of the objective of the recommendations.

6.2.1. Recommendations for medical practitioners

This research has discovered that the elimination of cardio-metabolic syndrome factors will improve the health status of the Bantu African from the Kinshasa Hinterland, DRC. Early medical detection and attention are required to prevent other opportunistic attacks. Putting people on proper diets to fight against hypertensive opportunistic symptoms, medical attention directed to advising attendants against obesity, and consumption of too much

A comparative analysis based on gender draws one's attention to males more than to females. Cardio-metabolic risk factors were more significantly pronounced among the males than among the females. A detailed understanding points at the idea that males are more predisposed to cardio-metabolic risk factors than females. The elimination of cardio-metabolic opportunistic symptomatic characteristics will definitely reduce the probability of vulnerability of development of cardiovascular attacks. The data analysis revealed that data based on the presence and absence of cardio-metabolic factors had completely different results. The following recommendations can be made from the analysis (Chapter 5) and conclusion (Chapter 6 (6.1)) of the abstract. The overall observation from this study is that this research has opportunities for further enhancement a function of further assessment. There is dire need for inclusion of more variables in the data. Variables such as smoking habits, alcohol consumption habits, vegetable consumption, etc. will play a positive role in future research.

6.2.2. Recommendation for Statisticians

The researcher has observed that the two types of analyses namely: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are at the centre of this statistical approach. The two

steps complement each other. This means that a successful execution of EFA has a strong bearing on the outcome of CFA.

There is, however, a confounding factor here which is the issue surrounding the establishment of factors at the EFA stage. The issue of interest is the heavy loading of indicators on many factors. At the time of identifying the "real" factors (components), during the EFA, researchers get confused when an indicator loads on many factors. Research has to be undertaken to draw a clear distinction between factors which receive heavy loadings by one indicator.

Bibliography

(EGIR), The European Group for the Study of Insulin Resistance. (2002). Frequency of the WHO metabolic syndrome in European cohorts, and an alternative definition of an insulin resistance syndrome.

(WHO), World Health Organization. (2006). World Health Organization (WHO) and International Diabetes Foundation (IDF) Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycemia. Geneva: WHO.

Aaon, J. (2002). Update on managing hypercholesterolemia. *50*(8).

Akaike, H. (1987). Factor analysis and AIC. . *Psychometrika*.

Akeroyd, M. A. (2014). A factor analysis of the SSQ (Speech, Spatial, and Qualities of Hearing Scale). *International Journal of Audiology*.

Akinpelu, A. O. (2008). Overweight and Obesity: Does it occur in Nigerian adolescents in an urban community? *Int J Biomed Hlth Sci., 4*, 11–7.

Alberti, K. Z. (2006). A new world definition: a consensus statement from the international diabetes federation.

Alberti, K. Z. (2006). Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *23*(5), 469-80.

Amoah A. G., O. S. (2002). Diabetes in Ghana: A Community Based Prevalence Study in Greater Accra. Diabetes Research and Clinical Practice. *PubMed*, 197–205.

Anderson, A. S. (2014). Child obesity and its impact on the development of adolescent.

Anderson, J. a. (1984). Improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis.

Anderson, P. C. (2001). Factor analysis of the metabolic syndrome: obesity vs insulin resistance as the central abnormality. *International Journal of Obesity*, 1782 -1788.

Anna, l. (2011).

Atochin D.N., W. A.-W. (2007). The phosphorylation state of eNOS modulates vascular reactivity.

Baldé N.M., D. I.-B. (2007). Diabetes and impared fasting glucose in rural and urban populations in futa jallon.

Barrett, P. (2007). Structural equation modelling: Adjudging model fi t. Personality and Individual Differences. *42*.

Barry, M. (2007). Diabetes and impared fasting glucose in rural and urban populations in futa jallon. *Diabetes mtab 2007, 33*(114-120).

Bartholomew D.J., K. . (2011). *Latent Variable Models and Factor Analysis.* John Wiley & Sons.

Bayauli MP, B.-K. J.-K. (2014). Prehypertension, Hypertension and Associated Cardiovascular Risk Factors among Adult Congolese Urban Dwellers: Results of the Vitaraa Study. *World Journal of Cardiovascular Diseases, 4*, 390-398. doi:10.4236/wjcd.2014.48049

Bensellam, M. (2012). Glucose-induced 02 consumption activities hypoxia inducible factors 1 and 2 in rat insulin-secreting pancreatic beta-cells.

Bentler, P. M. (2006). EQS 6 structural equations program manual.

Berna, F. e. (2015). Akathisia: prevalence and risk factors in a community-dwelling sample of patients with schizophrenia. Results from the FACE-SZ dataset.

Berna, G. (2015). Nutrigenetics and nutrigenomics insights into diabetes etiopathogenesis.

Bernstein, I. &. (1989). *Factoring items and factoring due to item categorization. Psychological Bulletin, 105, 467-477.* (Vol. 105).

Bernstein, I. &. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality. *Psychological Bulletin , 105*, 467-477. .

Blankson, A. a. (2014). A Brief Report on the Factor Structure of the Cognitive Measures in the HRS/AHEAD Studies. doi:10.1155/2014/798514

Bloomfield, G. S. (2015). Heart failure in sub-saharan africa. *157-173*.

Bollen, K. A. (1989). Structural equations with latent variables.

Booth, P. (2015). Lack of exercise is a major cause of chronic diseases. *2*(2.2).

Brambilla, P. (2007). metabolic risk factor clustering estimation in children: to draw a line across pediatric metabolic syndrome. *33*(591-600).

Brinkley, H. (2012). total and abdominal adiposity are associated with inflammation in older adults using a factor analysis approach. *67*(10).

Brockerhoff, M. (2000). Inequality of child mortality among ethnic groups in sub-sahsran africa. . *99*(136).

Brown, T. A. (2006). Confi rmatory factor analysis for applied research. .

Bryant, F. (1996). *Principal componnet analysis and exploratory and confirmatory factor analysis. in reading and understanding multivariate analysis.*

Buchanan, T. (2012). Gestational diabetes mellitus: RIsks and management during and after pregnancy.

Byrne, B. (1998). Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.

Byrne, B. (2001). Structural Equation Modelling with Amos, EQS,and LISREL: Comparative Approaches to Testing for the Factorial Validity of a Measuring Instrument: A two-study Analysis. *Journal of Business Ethics, 46*, 127-141.

Byrne, B. (2001). Structural equation modelling with AMOS:Basic concepts, applications, and programming.

Carty, C. B. (2014). Comparative Analysis of Metabolic Syndrome Components in over 15,000 African Americans Identifies Pleiotropic Variants: Results from the PAGE Study. *Circulation Cardiovascular genetics, 7*(4), 505-513. doi: 10.1161/CIRCGENETICS.113.000386

Cattell, R. (1966). The meaning and strategic use of factor analysis. Handbook of multivariate experimental psychology.

Cattell, R. (1973). A recent higher-order factor analysis of the Cattell, Comrey and Eysenck.

Cattell, R. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences. New York: .* New York.

Cattell, R. (2013). The scree plot test for the number of factors. multivariate behavioural research. *4*(10A).

Chackrewarthy S., G. D. (2013). A Comparison between Revised NCEP ATP III and IDF Definitions in Diagnosing Metabolic Syndrome in an Urban Sri Lankan Population: The Ragama Health Study.

Child, D. (2006). *The essentials of factor analysis. (3rd ed.). New York, NY: Continuum International Publishing Group.* New York.

Chirinos DA, M.-L. J. (2014). Metabolic Syndrome As an Underlying Disease Entity and Its Relationship to Subclinical Atherosclerosis in Andean Hispanics. . *Metabolic Syndrome and Related Disorders., 12*(1), 49-55. doi:10.1089/met.2013

Chukwuonye II, C. A. (2013). Prevalence of overweight and obesity in adult Nigerians - a systematic review.Diabetes Metab Syndr Obes6:43-47. *Family Med Prim Care 2: , 6*, 20-26.

Comrey AL, L. H. (1992). A First Course in Factor Analysis.

Corp., I. (2012). IBM SPSS Statistics for Windows (release 21.0) [Computer software]. Armonk,.

Costello, A. &. (2005). *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. Practical Assessment, Research and* (Vol. 10).

Crawford, A. V. (2010). Evaluation of parallel analysis methods for determining the number of factors. Educational and Psychological Measurement. *70*, 885–901. doi:10.1177/001316441037933

Creaton, M. I. (2015). Diabetes screening among immgrants: Apopulation-based urban Cohort study. *35*(4).

Davis, W. (1993). FCI rule of identification for confirmatory facotr analysis: A general sufficin=ent condition. *21*(4).

DeCoster J, S. E. (1998). Confirmatory factor analysis.

DeCoster, J. (1998). *Overview of Factor Analysis.*

Deurenberg, P. D.-Y. (2002). Validity of total and segmental impedance measurements for prediction of body composition across ethnic population groups. *56*(3), 214-220.

Deurenberg-Yap M, S. G. (2000). The paradox of low body mass index and high body fat percent among Chinese, Malays and Indians in Singapore. Int. J. Obes. Relat. Metab. Disord. *MEDLINE, 24*, 1011-1017. .

Diana, A. (2014). metabolic syndrome related disorder. *12*(1).

Dimmeler S., F. I. (1999). Activation of nitric oxide synthase in endothelial cells by Akt-dependent phosphorylation. Nature. *399*, 601–605.

DiStefano, C. (2002). The Impact of Categorization With Confirmatory Factor Analysis, Structural Equation Modeling: A Multidisciplinary Journal. *9*(3).

Elbagir, M. (1996). A population-based study of the prevalence of diabetes and impared glucose tolerance in adults in Northern sudan. *19*.

Esteghamati, A. (2016). Clustering of metabolic components in the middle eastern diabetic and non-diabetic population. *36*.

Fabrigar, L. R. (1999). Evaluating the use of exploratory factor analysis in psychological research psychological methods. Psychological Methods. *4*, 272–299. doi:10.1037/1082-989X.4.3.272

Fan, J. L. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *J. R. Stat. Soc. Ser. B.*.

Farook VS, R. L. (n.d.). Metabolites as novel biomarkers for childhood obesity-related traits in Mexican-American children.

Farook VS1, P. S. (2012). Metabolic syndrome is linked to chromosome 7q21 and associated with genetic variants in CD36 and GNAT3 in Mexican Americans. *PMC*. doi:10.1038/oby.2012.74

Farook, V. (n.d.). Comprehensive assessment of metabolic syndrome among rural Bangladeshi women.

Ferguson, E. (2009). A taxometric analysis of health anxiety. *Psuchological Medicine, 39*, 277-286.

Fezeu, L. (2007). Metabolic syndrome in sub-Saharan Africa setting: Central Obesity may be the key determinant.

Field, A. (2000). Discovering Statistics using SPSS for Windows.

Field, A. (2009.). Understand and interpret the structure of the set of variables.

Fighting for Intelligence A Brief Overview of the Academic Work of John L. Horn. Multivariate behavioral research . (n.d.).

Franke, D. J. (2015). Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. Nat Methods. *12*(5), 419-422.

Franke, L. C. (2015). Factor Analysis of Persistent Postconcussive Symptoms Within a Military Sample With Blast Exposure. *Journal of Head Trauma Rehabilitation, 30*(1). doi:10.1097/HTR.0000000000000042

Furfey, P. H. (1937). A criticism of factor analysis as a technique of social research. American Sociological Review. *2*, 178–186. doi:10.2307/2083471

Ghosh, A. (2005). Factor analysis of metabolic syndrome among the middle-aged bangalee Hindu men in Calcutta, India.

Gimbrone M.A., J. T.-C. (2000). Endothelial dysfunction, hemodynamic forces, and atherogenesis. *Ann NY Acad Sci. , 902*, 230–239; discussion 239–240. .

GM, R. (2006). The metabolic syndrome: is this diagnosis necessary? *Am J Clin Nutr. , 83*, 1237–1247 .

Gorsuch, R. (1983). *Factor Analysis, second edition,:.* Hillsdale: Lawrence Erlbaum Associates Inc.

Graham B, G. J. (2003). *Principles and Practice of structural Equation Modelling.* The Guilford Press.

Graham, G. &. (2003). Confirmatory factor analysis.

Graham, J. M. (2003). Consequences of not interpreting structure coeffi cients in published CFA research: A reminder. Structural Equation Modeling. *10*, 142–153. doi:10.1207/S15328007SEM1001_7

Gray, R. (2010). Risk factor clustering in insulin resistance syndrome.

Gray, R. (2010). Risk factor clustering in insulin resistance syndrome. The strong heart study.

Grice, J. W. (2001). Computing and evaluating factor scores. Psychological Methods. *6*, 430–450. doi:10.1037/1082-989X.6.4.430

Grundy SM, C. J. (2005). *Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement.*

Grundy SM, C. J., Association, A. H., & National Heart, L. a. (2005). Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. *112*(17).

Gu, D. (2005). InterASIS Collaborative Group: Prevalence of the metabolic syndrome and overweight among adults in china.

Gupta, S. C. (2000). *Fundamentals of Mathematical Statistics (A modern Approach).*

Gurka, M. ,. (2012). A confirmatory factor analysis of the metabolic syndrome in adolescents: an examination of sex and racial/ethnic differences. *BioMedCentral*.

Gurka, M. (2012). A confirmatory factor analysis of the metabolic syndrome in adolescents: an examination of sex and racial/ethnic differences.

Gyakobo M, A. A.-M.-A. (2012). Prevalence of the metabolic syndrome in a rural population in Ghana.. doi:10.1186/1472-6823-12-25

Hair J, B. W. (2006). Multivariate Data Analysis. *NJ*.

Hall V, T. R. (2011). Diabetes in Sub Saharan Africa 1999-2011: Epidemiology and public health implications. a systematic review.

Haller, H. (1977). Epidemiology and associated risk factors of hyperlipoproteinemia. *Jemdsa, 32*(8), 124 - 128.

Harman, H. (1976). Modern Factor Analysis.

Harman, H. (1976). Modern factor analysis (3rd ed. revised).

Heck, R. H. (2008). An introduction to multilevel modeling techniques.

Heiss G, M. M. (2014). Prevalence of Metabolic Syndrome among the Hispanics/Latinos of Diverse Background: The Hispanic Community Health Study/Study of the Latinos. *Diabetes Care*. doi:10.2337/dc 13-2505

Heiss G, S. M. (2014). Prevalence of metabolic syndrome among Hispanics/Latinos of diverse background: the Hispanic ... *37*(8), 2391-9. doi:10.2337/dc13-2505

HF, B. (2005). Ethical decision Making in the public Accounting Profession: An Extension of Ajzen's Theory of Planned Behaviour. *Journal Of Busiee Ethics, 61*, 165-181.

Higashi, Y. (2012). Quantifying options for reducing coronary heart disease mortality by 2020.

Hilawe EH, Y. H. (2013). Differences by sex in the prevalence of diabetes mellitus, impaired fasting glycaemia and impaired glucose tolerance in sub-Saharan Africa: a systematic review and meta-Analysis. *Bull World Health Organ , 91*, 671–682D.

Hillsten S., M. C. (2006). Are "Ethical" or Socially responsible" Investment Socially responsible? *Journal of Business Ethics, 66*(4), 393-406.

Hoffmann E, H. E. (2008). A comparison of models describing the impact of moral decision making on investment decisions. *J. Bus. Ethics, 82*, 171-187.

Hofmann M, P. K. (2000). Humans develop obesity, dyslipidemia, and arterial hypertension.

Huang, P. (2005). Unraveling the links between diabetes, obesity, and cardiovascular disease. *96*, 1129–1131 .

Huang, P. (2009). A comprehensive definition for metabolic syndrome. Disease Models & Mechanisms. *2*(5-6), 231-237. doi:10.1242/dmm.001180

Huo D, T. L. (2013). Association of blood pressure with development of metabolic syndrome components: a five-year Retrospective Cohort study in Beijing. doi:10.1186/1471-2458-13-912

Huo D, W. W. (2013). Evaluation of two single-factor models of metabolic syndrome: a confirmatory factor analysis for an adult population in Beijing. . *Lipids in Health and Disease, 12*, 61. doi:10.1186/1476-511X-12-61.

Hurley E.H, S. T. (1997). Exploratory and con®rmatory factor analysis:guidelines, issues, and alternatives. *JOURNAL OF ORGANIZATIONAL BEHAVIOR*, 667-683 .

Hurley, A. E. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behaviour, 18*, 667-683.

Iloh, G. U. (2015). Obesity and its Cardio-Metabolic co-morbidities among adult Nigerians in a primary care clinic of a tertiary hospital in South Estern Nigeria.

Jackman, S. (2002). *Notes on the Statistics of Factor Analysis.*

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some supsupport for the N:q hypothesis. Structural Equation Modeling,. *10*, 128-141. doi:10.1207/S15328007SEM1001_6

Jackson, J. (1980). Principal components and factor analysis. I- Principal components. *Journal of Quality Technology*.

Jesmin, S. I. (2012). Comprehensive assessment of metabolic syndrome among rural Bangladeshi women. doi:10.1186/1471-2458-12-49

Jolliffe, I. (1986). Principal Component. Analysis,. Second Edition.

Jonk, A. H. (2007). Microvascular dysfunction in obesity: a potential mechanism in the pathogenesis of obesity-associated insulin resistance and hypertension. *Physiology (Bethesda), 22*, 252-260.

Jöreskog, K. &. (2012). LISREL 9.1 for Windows [Computer software]. Scientifi c Software International.

Joreskog, K. (1970). Simultaneous Factor Analysis in Several Populations.

Julio, C. J. (2005). Prevalence of metabolic syndrome in an elderly community. comparison between three diagnostic methods.

Kaduka, L. K. (2012). Prevalence of Metabolic Syndrome Among an Urban Population in Kenya. Diabetes Care. *Diabetes Care, 35*(4), 887-893. doi:10.2337/dc11-0537.

Kahn R, B. J. (2005). The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *American Diabetes Association and European Association for the Study of Diabetes*, 17.

Kahn R1, B. J., Association, A. D., & Diabetes., E. A. (2005). The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. . *PubMed*.

Kaiser, H. (1958, September). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*(3), pp 187-200.

Kaiser, H. a. (1985). *Alpha factor analysis.* Psychometrika.

Kapetanovic, S. (2014). Paediatric HIV AIDS COhort Study Team biomarkers and neurodevelopment in perinatally HIV-infected or exposed youth. *28*(3).

Kapoor, V. a. (2000). *Fundamentals of Mathematical Statistics.*

Kasiam, L. (2009). Classification and dramatic epidemic of diabetes mellitus in Kinshasa Hinterland: the prominent role of type 2 diabetes and lifestyle changes among Africans. *18*(3).

kasiam, L. J. (2008). Prevalence and risk factors of diabetes mellitus in kinshasa Hinterland.

Kasiam, L. (n.d.). WHO metabolic syndrome in European cohorts, and an alternative definition of a World Health Organization (WHO) and International Diabetes Foundation (IDF).

Kaufman, A. S. (1983). K-ABC administration and scoring manual. Circle Pines. *American Guidance Service*.

Keith, T. Z. (1985). Questioning the K-ABC: What does it measure? School Psychology Review. *4*, 9-20.

Kelliny, C. (2008). Metabolic sundrome according to differnt definitions in a rapidly developing country of hte African region.

Kengne AP, N. L. (2012). A systematic overview of prospective cohort studies of cardiovascular disease in sub-Saharan Africa. *Cardiovascular Journal of Africa, 23*(8), 469–470.

Kenny, D. A. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), The handbook of social psychology (Vol. 1, 4th ed., pp. . *1*, 233–265.

Kershaw, E. n. (2004). Adipose tissue as an endocrine organ. *J Clin Endocrinol Metab, 89*, 2548–2556. doi:10.1210/jc.2004-0395

Khader, Y. (2011). Factor analysis of cardio-metabolic risk factors clustering in children and adolescents.

Kieffer, K. (1999). An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. Research in the Schools. *6*(2), 75-92.

Kieffer, T. a. (1999). 1999 Dec;20(6):. The glucagon-like peptides. *20*(6), 876-913.

Kim E., L. N. (2010). Change in the rate of metabolic syndrome in patients with schizophrenia and bipolar disorder in the course of treatment. *12*, 31-31.

Kim J.A., M. M. (2006). Reciprocal relationships between insulin resistance and endothelial dysfunction: molecular and pathophysiological mechanisms. *113*, 1888–1904 .

Kim, S. K. (2003, March). Factor analysis of the metabolic syndrome among elderly Koreans - The South-west Seoul Study. *20*(2). doi:10.1046/j.1464-5491.2003.00890.x

Kline, R. (1994). *An Easy Guide to Factor Analysis.*

Kline, R. (2011). Principles and Practice of Structural Equation Modelling.

Kline, R. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher & C. Schatsschneider (Eds.), Applied quantitative analysis in the social sciences. 171-207.

Knafl, G. J., & Grey, M. (2007). Factor analysis model evaluation through likelihood cross validation. *Stat Methods Med Res., 16*(2), 77–102.

Kolenikov, S. (2009). Confirmatory factor analysis using confa. *The Stata Journal, 9*(3), pp. 329–373.

Krumsiek J, M. K. (2015). Gender-Specific pathway differences in human serum metabolome.

Krumsiek J, S. V. (2015). Multi-omic signature of body weight change: results from a population-based cohort study. doi:10.1186/s12916-015-0282-y

Kurmsiek, J. (2015). Gender-specific pathway differnces in the human serum metabolome.

Lawson M. A., K. K. (2011). Intracerebroventricular administration of HIV-1 Tat induces brain cytokine and indoleamine 2,3-dioxygenase expression: a possible mechanism for AIDS comorbid depression. Brain Behav. Immun. . *25*, 1569–1575. doi: 10.1016/j.bbi.2011.05.006

Lawson, D. (2015). Inference of Population Structure Using Dense Haplotype Data.

Lawson, T. K. (2012). Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Current Opinion in Biotechnology. *23*, 215–220.

Lee, J. (2007). Effective cut-off values of waist circumference to detect factors of metabolic syndrome in Japanese men and women.

Little, T. D. (2002). To parcel or not to parparcel: Exploring the question, weighing the merits. Structural Equation Modeling. *9*, 151-173. doi:10.1207/S15328007SEM0902_1

Liu, Q. Z. (2012). Effects of Head-Down Bed Rest on the Executive Functions and Emotional Response. *7*(12). doi:10.1371/journal.pone.0052160

Liu, S. (2015). Endothelial Nitric-Oxide Synthase (eNOS) Is Activated through G-Protein-Coupled Receptor Kinase-Interacting Protein 1 (GIT1) Tyrosine Phosphorylation and Src Protein. The Journal of Biological Chemistry 289.26 .

Lloyd-Jones, D. (2010). Executive summary: heart disease and stroke statistics--2010 update: a report from the American Heart Association.

Longo-Mbenza, B. (1999). Low birth weight and risk of hypertension in African school children.

Longo-Mbenza, B. (2007). Socioeconomic Pathology As a Cause of Regional Differences in the Prevalence of Metabolic Syndrome and Pregnancy-Induced Hypertension. *60*.

Longo-Mbenza, B. (2010). Low birth weight, metabolic syndrome and their associations with the global crisis of 1930 - 1945, rapidly growing economy and coronary heart disease in Central Africa.

Longo-Mbenza, B. (2010). Metabolic syndrome, aging, physical inactivity, and incidence of type 2 diabetes in general African population.

Longo-Mbenza, B. (2011). The metabolic syndrome in a Congolese population and its implications for metabolic syndrome definitions.

Ma, R. C. (2013). Type 2 Diabetes in East Asians: Similarities and Differences with Populations in Europe and the United States.

Maat, S. A. (2015). Confirmatory Factor Analysis of Learning Environment Instrument among High performance School Studies.

MacCallum R.C Widaman, K. P. (2011). *Sample Size in Factor Analysis: The Role of Model Error.*

MacCallum, R. (1983). A comparison of factor analysis programs in SPSS, BMDP, and SAS. *48*.

MacCallum, R. B. (n.d.). *"Power Analysis and Determination of Sample Size for Covariance Structure modelling," Psychological Methods, 1 (2), 130-49.*

MacCallum, R. B. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modelling,. *1*(2), 130-49.

MacCallum, R. C. (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological Methods. *1*, 130-149. doi:10.1037//1082-989X.1.2.130

MacCallum, R. C. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.

Magalie Thierry, B. P.-G.-N. (2014). Metabolic Syndrome Triggered by High-Fructose Diet Favors Choroidal Neovascularization and Impairs Retinal Light Sensitivity in the Rat. doi:10.1371/journal.pone.0112450

Magnussen, G. (2015). Pediatric Metabolic Syndrome Predicts Adulthood Metabolic Syndrome, Subclinical Atherosclerosis, and Type 2 Diabetes Mellitus – But Is No Better Than Body Mass Index Alone: The Bogalusa Heart Study and the Cardiovascular Risk in Young Finns Study.

Martinez-Vizcaino V, M. M. (2010). Validity of a single-factor model underlying the metabolic syndrome in children:a confirmatory factor analysis. *Diabetes Care , 33*, 1370–1372.

Martínez-Vizcaino, V. O.-M. (2010). Stability of the factorial structure of metabolic syndrome from childhood to adolescence: a 6-year follow-up study. doi:10.1186/1475-2840-10-81

Matsha, T. (2015). Optimal Waist-to-Height Ratio Values for Cardiometabolic Risk Screening in an Ethnically Diverse Sample of South African Urban and Rural School Boys and Girls.

McArdle, J. (2015). Fighting for Intelligence A Brief Overview of the Academic Work of John L. Horn. Multivariate behavioral research .

McCallery, J. (2012). Factor Structure Underlying Components of Allostatic Load.

McDonald, R. (1982). A note on the investigation of local and global identifiability. *47*(1).

Meade, A. W. (2004). A comparison of item response theory and confi rmatory factor analytic methodologies for establishing measurement equivalence/invariance Organizational Research Methods . *7*, 361–388. doi:10.1177/1094428104268027

Meigs JB. (2004). Metabolic syndrome: in search of a clinical role. Diabetes Care 27,. *Diabetes Care, 27*, 2761–2763.

Meigs, J. (2000). Syndrome X? Multiple Metabolic Syndrome? A Syndrome At All? Factor Analysis Reveals Patterns in …. *American Journal of Epidemiology*.

Millsap, R. E. (2007). Structural equation modeling made diffi cult. Personality and Individual Differences. *42*, 875–881 . doi:10.1016/j.paid.2006.09.021

Misra A, C. P. (2009). Consensus Statement for Diagnosis of Obesity, Abdominal Obesity and the Metabolic Syndrome for Asian Indians and Recommendations for Physical Activity, Medical and Surgical Management. *J Assoc Physicians India , 57*, 163–170.

Misra, A. (2013). Obesity and Dyslipidemia in South Asians.

Mithun, D. (2010). Factor analysis of risk variables associated with metabolic syndrome in adult Asian Indians.

Motala, A. (2003). High incidence of type 2 diabetes mellitus in South African Indians:10 years of study.

Mpembe, B. L.-K. (2015). Left Ventricular Structure, Geometry and Systolic Function among Hypertensive Black Patients with Reduced Kidney Function. *World Journal of Cardiovascular Diseases, 5*, 287-295. doi:10.4236/wjcd.2015.510032

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. Multivariate Behavioral Research,. *22*, 267–305. doi: 10.1207/s15327906mbr2203_3

Mulaik, S. A. (1992). Theme issue on principal components analysis. Multivariate Behavioral Research. *27*(3).

Mulaik, S. A. (2009). Foundations of factor analysis (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Muniyappa R, L. S. (2008). Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage. *Am J Physiol Endocrinol Metab :, 294*, E15–E26.

Muniyappa, R. (2015). An Integrated View of Insulin Resistance and Endothelial Dysfunction. Endocrinology and metabolism clinics of North America .

Muthen, B. a. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British journal of mathematical & statistical psychology*.

Muthén, L. K. (1998-2012). Mplus user's guide (7th ed.). Los Angeles, CA:.

Nasila, J. S. (2013). Assessing clustering of metabolic syndrome components available at primary care for Bantu Africans using factor analysis in the general population. *BMC*, 10.

Ngondo, A. (1984). Demographic perspectives of Zaïre during 1984–1999. Cedas 1984, 1:31-32. .

Nunnally, J. C. (1994). Psychometric theory (3rd ed.). .

O'Connor, B. (. (2011). Cautions regarding item-level factor analyses. Retrieved January 27, 2011, fromhttps://people.ok.ubc.ca/brioconn/nfactors/itemanalysis.html.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments & Computers. *32* , 396–402. doi:10.3758/BF03200807

Ofotokun I, S. A. (2012). A switch in therapy to a reverse transcriptase inhibitor sparing combination of lopinavir/ritonavir and raltegravir in virologically suppressed HIV-infected patients: a pilot randomized trial to assess efficacy and safety profile: the KITE study. *AIDS Research Hum Retroviruses, 28*, 1192-206.

Okafor, C. (2012). The Metabolic Syndrome in Africa: Current Trends. Indian Journal of Endocrinology and Metabolism.

Olinto, M. T. (2015). Major Dietary Patterns and Cardiovascular Risk Factors among Young Brazilian Adults.

Ong, K.-L. (2014). The Relationship between Insulin Resistance and Vascular Calcification in Coronary Arteries, and the Thoracic and Abdominal Aorta.

Papanastasiou, E. (2013). The prevalence and mechanisms of metabolic syndrome in schizophrenia: a review. *Therapeutic Advances in Psychopharmacology, 3*(1), 33-51.

Pedhazur, E. J. (1991). *Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.* NJ: Hilsdale.

Pervanidou, P. (2007). Assessment of metabolic profile in a clinical setting. .

Prasad, D. S. (2011). Abdominal Obesity, an Independent Cardiovascular Risk Factor in Indian Subcontinent: A Clinico Epidemiological Evidence Summary.

Program., N. C. (2002). Update on managing hypercholesterolemia. The new NCEP guidelines. *AAOHN J. 2002 , 50*(8), 360-4.

Purdon, P. (2015). Confirmatory factor analysis and structural equation modelling.

Rafique, T. Z. (2015). Onset of Metabolic Syndrome: A Short Review on the Current Risk Towards Bangladeshi Women. *American Journal of BioScience, 3*(3), 114-116. doi:10.11648/j.ajbio.20150303.16

Rampal S, M. S. (2012). Ethnic Differences in the Prevalence of Metabolic Syndrome.

Reaven, G. (2004). Insulin resistance and its consequences. In Diabetes Mellitus: A Fundamental and Clinical Text.

Regitz-Zagrosek, V. (2012). Sex and gender differences in pharmacology. .

Revicki, D. A. (2014). Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. Quality of Life Research, 23(1). *23*(1), 245-255. doi:10.1007/s11136-013-0467-9

Riccardo, D. (2015). Major factors for facilitating change in behavioural strategies to reduce obesity. *6*(101-110).

Rice J.A. (2010). *Mathamatical Statisitcs.*

Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. Multivariate Behavioral Research.

Roberts, C. K. (2013). Metabolic Syndrome and Insulin Resistance: Underlying Causes and Modification by Exercise Training. Comprehensive Physiology, . *3*(1), 1–58. doi: 10.1002/cphy.c110062

Rodgers, R. a. (1995). Factor analysis of spatiotemporal and ethological measures in the murine elevated plus-maze test of …. Pharmacol Biochem Behav.

Rodriguez KF1, U. E.-M. (2015). Effects of in Utero Exposure to Arsenic during the Second Half of Gestation on Reproductive End Points and Metabolic Parameters in Female CD-1 Mice. *Environ Health Perspective*.

Rodriguez, C. (2014). Status of Cardiovascular Disease and Stroke in Hispanics/Latinos in the United States: A Science Advisory From the American Heart Association.

Roger, V. L. (2015). Heart Disease and Stroke Statistics—2011 Update: A Report From the American Heart Association.

Ronald H. Heck, S. L. (2000). *An Introduction to Multilevel modelling Techniques.* Lawrence Erlbaum Associates.

RP, M. (1985). Factor Analysis and related methods.

RUMMEL, R. (1988). Applied Factor Analysis. books.google.com. [Cited by 352] (19.22/year).

Russell, D. (2002). .In search of underlying dimensions: the use (and abuse) of factor analysis in personality and social psychology bulletin.

Russo, I. (2012). The Prothrombotic Tendency in Metabolic Syndrome: Focus on the Potential Mechanisms Involved in Impaired Haemostasis and Fibrinolytic Balance.

S, S. (2003). Socially responsible investing in the U S. *J. Bus. Ethics, 43*, 189-194.

Samaras K, G. S. (2008, November 13). Proinflammatory markers, insulin sensitivity, and cardiometabolic risk factors in treated HIV infection. *PubMed, 1*, 53-59. doi:10.1038/oby.2008.500

Santor, D. e. (2011). An Overview of Confirmatory Factor Analysis and Item Response Analysis Applied to Instruments to Evaluate Primary Healthcare. *7*(Special).

SAS Institute, I. (2012). SAS/STAT (release 12.1) [Computer software]. Cary, NC: Author.

Satoh, T. (2014). Rho GTPases in Insulin-Stimulated Glucose Uptake. Small GTPases .

Satoh, T. (2014). Rho GTPases in Insulin-Stimulated Glucose Uptake. Small GTPases .

Schmitt, T. A. (2011). Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis. *Journal of Psychoeducational Assessment, 29*(4), 304–321.

Semenkovich, C. (2006). Insulin resistance and atherosclerosis. *Journal of Clinical Investigations, 116*, 1813–1822.

Shafer WE. (2008). Ethical Climate in China CPA firms. Account.org. and. *J. Bus. Ethics, 33*, 825-835.

Shah, R. &. (2006). Use of structural equation modeling in operations managementLooking back and forward. Journal of Operations Management. *24*, 148–169. doi:10.1016/j.jom.2005.05.001

Shai I, J. R. (2006). Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. *Diabetes Care, 29*(7), 1585-90.

Shen, B. e. (2002). Are Metabolic Risk Factors One Unified Syndrome? Modeling the Structure of the Metabolic Syndrome X. *American Journal of Epidemiology, 157*.

Sheppard, A. (1996). The sequence of factor analysis and cluster analysis: differences in segmentation and dimensionality through the use of raw and factor scores.

Silvia, E. S. (1988). Some factors affecting the success of specifi cation searches in covariance structure modeling. Multivariate Behavioral Research. *23*(297–326). doi:10.1207/s15327906mbr2303_2

Simin Liu, M. S. (20125). Analysis of Metabolic Syndrome Components in >15 000 African Americans Identifies Pleiotropic Variants.

Sinclair, E., & Alexson, J. (1986). Factor Analysis and Discriminant Analysis of Psychoeducational Report Contents. *Journal of School Psychology, 24*(4), 363-371. doi:http://dx.doi.org/10.1016/0022-4405(86)90024-5

Smits, M. (2013). Adipocytokines as Features of the Metabolic Syndrome Determined Using Confirmatory Factor Analysis. Annals of Epidemiology.

Smitts, M. S.-K. (2015). Low consumption of fruits and dairy foods is associated with metabolic syndrome in Korean adults from outpatient clinics in and near Seoul. *9*(5).

Sörbom, J. e. (2004). Structural Equation Modelling.

Spearman, C. (1904a). The proof and measurement of the association between two things. American. *Journal of Psychology, 15*, 72-101. doi:10.2307/1412159

Spearman, C. (1904b). “General intelligence,” objectively determined and measured. . *American Journal of Psychology, 15*, 201–293. doi:10.2307/1412107

Sternberg, R. (1990). The geographic metaphor. In Metaphors of mind: conceptions of the nature of intelligence. Edited by Sternberg RJ. New York: Cambridge.

Stills, D. (1989). International encyclopedia of the social sciences: biographical supplement. .

Sullivan PJ., L.-M. K. (2015). Confirmatory Factor Analysis of the Youth Experiences Survey for Sport (YES-S).

Tabachnick, B. G. (2001). *Using multivariate statistics.* New York: Allyn and Bacon.

Tangvarasittichai, S. (2015). Oxidative stress, insulin resistance, dyslipidemia and type 2 diabetes mellitus. *World Journal Diabetes , 6*, 456-480.

Thompson, B. (2004). Exploratory and confirmatory factor analysis. *American Psychological Association.*

Thomson, G. H. (1920). General versus group factors in mental activities. *Psychological Review*(27), 173–190 . doi:10.1037/h0068628

Tucker, L. a. (1997). *Exploratory Factor Analysis.*

Ugwuja, E. O. (2013). Overweight and obesity, lipid profile and athegenic indicies among civil servants in Abakaliki, South Eastern Nigeria. *Annals of medical and health sciences research, 3*(1).

Urbina, E. M. (2011). Insulin resistance and arterial stiffness in healthy adolescents and young adults.

USAID/WHO/UNICEFV. (2005). Democratic Republic of Congo: National survey of children and women.

Van Prooijen, J. &. (2001). Confirmatory analysis of exploratively obtained factor structures. Educational and Psychological Measurement, 61, 777–792. *61*, 777-792. doi:10.1177/00131640121971518

Velicer, W. a. (1988). The relationship of sample size to the stability of component patterns: A simulation study. *Psychological Bulletin, 103*, 265-275.

Velicer, W. F. (1990a). Component analysis versus common factor analysis Some issues in selecting an appropriate procedure. Multivariate Behavioral Research. *25*, 1–28. doi:10.1207/s15327906mbr2501_1

Von Hippel, P. P. (2011). Elementary school-based obesity prevention intervention effect on waist ... *42*, 235–45.

Wang M, A. C. (2013). Factor structure of the CES-D and measurement invariance across gender in Mainland Chinese adolescents. *Journal of Clinical Psycholofy*. doi:10.1002/jclp.21978

Watson D, a. T. (2006). Exploratory and confirmatory factor analysis.

Watson, R. &. (2006). Use of factor analysis in Journal of Advanced Nursing: Literature review. Journal of Advanced Nursing,. *55*. doi:10.1111/j.1365-2648.2006.03915.x

Weissglas-Volkov, D. a. (2010). Genetic causes of high and low serum cholesterol. *Journal of Lipid Research., 51*, 2032–2057.

Westley, R. L. (2013). A Twenty-First Century Cancer Epidemic Caused by Obesity. The Involvement of Insulin, Diabetes, and Insulin-Like Growth Factors.

WHO Expert Committee. (1995).

Widaman, K. F. (1993). Common factor analysis versus principal components analysis: “Differential bias in representing model parameters?" Multivariate Behavioral Research 28: 263-311.

Williams, B. B. (2012). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine , 8*(3).

Woolston, A. (2012). A Comparison of Different Approaches to Unravel the Latent Structure within Metabolic Syndrome.

World Health Organization (WHO). (2011). The metabolic syndrome in a Congolese population and its implications for metabolic syndrome definitions. Diabetes & metabolic syndrome. Clinical Research & Reviews 2011, 5:17–24. The Prevalence of the Metabolic Syndrome in Korean Adults:Comparison of WHO.

Wu, C. L. (2006). Factor analysis of metabolic syndrome using direct measurement of insulin resistance in Chinese with different degrees of glucose tolerance. 336 - 343.

Zhang, W. a. (2013). Thresholding Principal Orthogonal Complements.” Journal of the Royal Statistical Society.

Zimmet, P. (2007). IDF Consensus Group: The metabolic syndrome in children and adolescents.

Appendix

A.1 Questionnaire used for data collection

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format	Missing Values
Num	*1*	*Number*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
ORIGIN	*2*	*Rural or Urban*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
rural	*3*	*Cases or controls under rural*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
urbain	*4*	*Cases or controls under urban*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
AGE	*10*	*Age in years*	*Scale*	*Input*	*8*	*Right*	*F3*	*F3*	
AGEGRP1	*11*	*Age groups1*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
AGEGRP2	*12*	*Age group2*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
AGEGRP3	*13*	*Age group3*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
AGEGRP4	*14*	*Age group4*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
GENDER	*15*	*Gender*	*Nominal*	*Input*	*8*	*Right*	*F1*	*F1*	
q3	*16*	*Region of origin*	*Ordinal*	*Input*	*12*	*Right*	*F2.1*	*F2.1*	*99.0*
q4	*17*	*Head of household*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	

q5	*18*	*Religion*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
q6	*19*	*Marital status*	*Nominal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q7	*20*	*Delivered/ given birth*	*Nominal*	*Input*	*8*	*Right*	*F1*	*F1*	*99*
q8	*21*	*macrosomia*	*Nominal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q9	*22*	*Ethnicity*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q10	*23*	*Origin of ethnicity*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q11	*24*	*Race*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
q12	*25*	*Worker*	*Nominal*	*Input*	*8*	*Right*	*F1*	*F1*	
q13	*26*	*Type of work*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
m1	*27*	*Works for more than 9 hours*	*Scale*	*Input*	*8*	*Right*	*F8.2*	*F8.2*	
MONTHLY.INCOME	*28*	*Monthly income*	*Ordinal*	*Input*	*12*	*Right*	*F2*	*F2*	*99*
r1	*29*	*Monthly income1*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
r2	*30*	*Monthly income2*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
r3	*31*	*Monthly income3*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
t1	*32*	*Private transport*	*Scale*	*Input*	*8*	*Right*	*F1*	*F1*	
q17	*33*	*Type of transport*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
i1	*34*	*Knowledge about the risk factors of DM*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	

q19	*35*	*Educational attainment*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
e1	*36*	*Educational attainment1*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
e2	*37*	*Educational attainment2*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
e3	*38*	*Educational attainment3*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
e4	*39*	*Educational attainment4*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
q20	*40*	*Number of meals per day*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
q21	*41*	*Have breakfast*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
g1	*42*	*Animal fat*	*Scale*	*Input*	*14*	*Right*	*F8*	*F8*	
q23	*43*	*Viscera*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
a1	*44*	*Alcohol intake*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
q25	*45*	*Duration of alcohol intake*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q26	*46*	*Type of alcohol consumed*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
q27	*47*	*Stopped drinking*	*Nominal*	*Input*	*8*	*Right*	*F1*	*F1*	

q28	*48*	*How long you stopped drinking?*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q29	*49*	*Quantity of alcohol consumed*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	*99*
q30	*50*	*Frequency of fruits and vegetable intake*	*Ordinal*	*Input*	*8*	*Right*	*F1*	*F1*	
s1	*51*	*Refined sugar intake*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
q32	*52*	*Cigarette smoking*	*Nominal*	*Input*	*8*	*Right*	*F1*	*F1*	
d1	*53*	*Physical activity practice*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*	
q37	*54*	*Diabetes knowledge*	*Nominal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q38	*55*	*Duration of diabetes in years*	*Scale*	*Input*	*8*	*Right*	*F8.2*	*F8.2*	
q38bis	*56*	*Categories of diabetes duration*	*Ordinal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
FHD	*57*	*Family history of diabetes*	*Nominal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*
q47	*58*	*History of mulnutrition in childhood*	*Nominal*	*Input*	*8*	*Right*	*F2*	*F2*	*99*

SBP	*59*	*SYSTOLIC BP*	*Scale*	*Input*	*8*	*Right*	*F3*	*F3*
DBP	*60*	*DIASTOLIC BP*	*Scale*	*Input*	*8*	*Right*	*F3*	*F3*
WEIGHT	*61*	*WEIGHT*	*Scale*	*Input*	*8*	*Right*	*F3*	*F3*
HEIGHT	*62*	*HEIGHT*	*Scale*	*Input*	*8*	*Right*	*F8.2*	*F8.2*
BMI	*63*	*BODY MI*	*Scale*	*Input*	*8*	*Right*	*F8.1*	*F8.1*
GLFASTING	*64*	*GLUCOSE FA*	*Scale*	*Input*	*9*	*Right*	*F8*	*F8*
WST.CIRCUM	*65*	*WAIST CIRC*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
BMICAT2	*66*	*BMI2*	*Scale*	*Input*	*16*	*Right*	*F8*	*F8*
BMICAT3	*67*	*BMI3*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
BMICAT4	*68*	*BMI4*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
WAIST.CIRCUM	*69*	*WAIST CIRCUMFERENCE*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
ABDOM_OBESITY.FEMALES	*70*	*ABDOMINAL OBESITY IN FEMALES*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
ABDOM_OBESITY.MALES	*71*	*ABDOMINAL OBESITY IN MALES*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
GFASTING	*72*	*GLUCOSE FASTING*	*Scale*	*Input*	*9*	*Right*	*F8*	*F8*
GLUCOSE1	*73*	*Glucose1*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
GLUCOSE2	*74*	*GLUCOSE2*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*

hjnd	*75*	*FASTING HYPERGLICEMIA*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
PLGLUCOSE	*76*	*POST LG*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
pl_glucose	*77*	*Plglucose 1*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
intogluc	*78*	*Fasting glucose intolerance*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
diab1	*79*	*DM/fasting glucose*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
diab2	*80*	*DM/post load glucose*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
o1	*81*	*Abdominal obesity in all*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
DIABETES	*82*	*ALL DIABETES*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
PRE_DIABETES	*83*	*PREDIABETES*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
GTD	*84*	*Glucose tolerance disorders*	*Scale*	*Input*	*8*	*Right*	*F8*	*F8*
HIGH_BP	*85*	*HIGH BLOOD PRESSURE*	*Scale*	*Input*	*10*	*Right*	*F8.2*	*F8.2*
HYPERTENSION	*86*	*HYPERTENSION*	*Scale*	*Input*	*14*	*Right*	*F8.2*	*F8.2*
ALLDM2	*87*	*All diabetics2*	*Scale*	*Input*	*10*	*Right*	*F8.2*	*F8.2*

CARDIOMETABOLICRISK	*88*	*CARDIO METABOLIC RISK*	*Nominal*	*Input*	*14*	*Right*	*F8.2*	*F8.2*
FAC1_1	*89*	*REGR factor score 1 for analysis 1*	*Scale*	*Input*	*13*	*Right*	*F11.5*	*F11.5*
FAC2_1	*90*	*REGR factor score 2 for analysis 1*	*Scale*	*Input*	*13*	*Right*	*F11.5*	*F11.5*
FAC3_1	*91*	*REGR factor score 3 for analysis 1*	*Scale*	*Input*	*13*	*Right*	*F11.5*	*F11.5*
FAC1_2	*92*	*REGR factor score 1 for analysis 2*	*Scale*	*Input*	*13*	*Right*	*F11.5*	*F11.5*
FAC2_2	*93*	*REGR factor score 2 for analysis 2*	*Scale*	*Input*	*13*	*Right*	*F11.5*	*F11.5*
SBP2	*94*	*SYSTOLIC BLOOD PRESSURE (Binned)*	*Ordinal*	*Input*	*10*	*Right*	*F5*	*F5*
DBP2	*95*	*DIASTOLIC BLOOD PRESSURE (Binned)*	*Ordinal*	*Input*	*10*	*Right*	*F5*	*F5*

DBP3	*96*	*DIASTOLIC BLOOD PRESSURE (Binned)*	*Ordinal*	*Input*	*10*	*Right*	*F5*	*F5*
AGE_1	*97*	*MA(AGE, 1,1)*	*Scale*	*Input*	*7*	*Right*	*F5.1*	*F5.1*

Variable Values

Value		Label
ORIGIN	1	*Rural*
	2	*Urban*
rural	1	*cas rural*
	2	*temoins rural*
urbain	1	*cas urbain*
	2	*temoins urbain*
casrural	1	*cas Kikwit*
	2	*cas villages*
temrural	1	*temoins kikwit*
	2	*temoins villages*
caskinsh	1	*Gombe case*
	2	*Kindele case*
temkinsh	1	*Gombe control*
	2	*Kindele control*
castem	0	*Controls*
	1	*Cases*
AGEGRP1	1	*<40 Years*

	2	*>= 40 years*
AGEGRP2	1	*<25 years*
	2	*25-45 years*
	3	*>45 years*
AGEGRP3	1	*< 15 years*
	2	*15-64 years*
	3	*>64 years*
AGEGRP4	1	*18-24 years*
	2	*25-34 years*
	3	*35-44 years*
	4	*45-54 years*
	5	*55-64 years*
	6	*65-74 years*
	7	*75 years +*
GENDER	1	*Male*
	2	*Female*
q3	1.0	*Kinshasa*
	2.0	*Bas Congo*
	3.0	*Equator*
	4.0	*Province Orientale*
	5.0	*Bandundu*
	6.0	*Kasai Occidental*
	7.0	*Kasai Oriental*
	8.0	*Kivu*
	9.0	*Katanga*
	10.0	*Aucune*
q4	1	*Self*
	2	*Father*
	3	*Mother*

	4	*Husband*
	5	*Spouse*
	6	*Friend*
	7	*Brother*
q5	1	*Catholic*
	2	*protestant*
	3	*kimbanguiste*
	4	*Muslim*
	5	*pentecostal*
	6	*Other*
	7	*None*
q6	1	*Yes*
	2	*No*
q7	1	*Yes*
	2	*No*
q8	1	*Yes*
	2	*No*
q9	1	*Congo*
	2	*Luba*
	3	*Ngala*
	4	*Swahili*
	5	*None*
q11	1	*Black*
	2	*White*
	3	*None*
q12	1	*Yes*
	2	*No*
q13	1	*Farmer*
	2	*House wife*

	3	*Metal worker*
	4	*Business man*
	5	*Civil Servant*
	6	*Senior executive*
	7	*Other*
m1	1.00	*Yes*
	2.00	*No*
MONTHLY.INCOME	1	*< 100$*
	2	*100$ - 200$*
	3	*201$ -600$*
	4	*> 600$*
r1	1	*< $100*
	2	*>= $ 100*
r2	1	*100-200*
	2	*Other*
r3	1	*< $ 200*
	2	*>= $ 200*
t1	1	*Yes*
	2	*No*
q17	1	*Walking*
	2	*Bicycle*
	3	*By car*
	4	*None*
i1	0	*No*
	1	*Yes*
q19	1	*Primary*
	2	*Professional*
	3	*Secondary*
	4	*University*

	5	*Illiterate*
e1	0	*Other*
	1	*1*
e2	0	*Other*
	1	*2*
e3	0	*Others*
	1	*3*
e4	0	*Others*
	1	*4*
q20	1	*< 1*
	2	*1*
	3	*2*
	4	*3*
	5	*>3*
q21	1	*Breakfast*
	2	*lunch*
	3	*dinner*
	4	*appetizer*
	6	*All the time*
	7	*desert*
	8	*biberonnage*
	9	*other*
g1	0	*Low frequency intake*
	1	*High intake*
q23	1	*Weekly*
	2	*Two week intervals*
	3	*Three week intervals*
	4	*4 week intervals*
	5	*Five or more week intervals*

	6	*Never*
a1	0	*No*
	1	*Yes*
q25	1	*<1 year*
	2	*1 - 4 years*
	3	*5 - 10 years*
	4	*> 10 years*
q26	1	*Beer*
	2	*Palm wine*
	3	*whisky*
	4	*Cane brew*
	5	*Local spirit*
	6	*Other*
q27	1	*Yes*
	2	*No*
q28	1	*< 1 years*
	2	*1 - 5 years*
	3	*6 - 10 years*
	4	*> 10 Years*
q29	1	*<1 bottle*
	2	*1 - 4 bottles*
	3	*5 - 11 bottles*
	4	*12 - 24 bottles*
	5	*> 24 bottles*
q30	1	*Daily*
	2	*2 - 5 days interval*
	3	*6 - 10 day interval*
	4	*>10 day interval*
s1	0	*Low intake*

	1	*High intake*
q32	1	*Yes*
	2	*No*
d1	0	*No*
	1	*Yes*
q37	1	*Yes*
	2	*No*
q38bis	1	*< 1 year*
	2	*1 - 2 years*
	3	*3 - 5 years*
	4	*6 - 10 years*
	5	*11 - 15 years*
	6	*>15 years*
FHD	1	*Yes*
	2	*No*
q47	1	*Yes*
	2	*No*
BMICAT2	1	*BMI <25 kg/m2*
	2	*BMI >=25 kg/m2*
BMICAT3	0	*BMI >=18.5 kg/m2*
	1	*BMI < 18.5 kg/m2*
BMICAT4	0	*BMI <30 kg/m2*
	1	*BMI >= 30kg/m2*
ABDOM_OBESITY.FEMALES	0	*WC <88 cm*
	1	*WC >=88 cm*
ABDOM_OBESITY.MALES	0	*WC< 102 cm*
	1	*WC >=102 cm*
GLUCOSE1	1	*<111*
	2	*111-125*

	3	*>125*
GLUCOSE2	1	*<101*
	2	*101-125*
	3	*>125*
hjnd	0	*Other*
	1	*101-125*
pl_glucose	1	*<140*
	2	*140-199*
	3	*>199*
intogluc	0	*Other*
	1	*140-199mg/dL*
diab1	0	*Other*
	1	*>125*
diab2	0	*Other*
	1	*>199mg/dL*
o1	0	*Absence*
	1	*Presence*
DIABETES	0	*Absence*
	1	*Presence*
PRE_DIABETES	0	*Other*
	1	*HJND et INTOGLU*
GTD	0	*Absence*
	1	*Presence*
CARDIOMETABOLICRISK	1.00	*Presence*
	2.00	*Absence*

A.2 Sample data

Table 106: Sample of the Data used in the Research.

SBP	DBP	WEIGHT	HEIGHT	BMI	GLFASTING	WST.CIRCUM
110	60	61	1.52	26.4	142	86
200	70	62	1.76	20.0	147	96
#NULL!	#NULL!	50	1.50	22.2	120	85
120	70	52	1.69	18.2	110	75
180	110	69	1.55	28.7	134	100
110	70	48	1.71	16.4	121	65
90	70	47	1.55	19.6	113	55
200	100	66	1.58	26.4	134	100
90	60	29	1.35	15.9	130	63
110	80	52	1.59	20.6	255	77
110	80	61	1.72	20.6	114	73
90	70	58	1.78	18.3	189	74
150	90	61	1.57	24.7	126	74
90	60	45	1.71	15.4	120	77
140	70	51	1.62	19.4	301	76
110	70	55	1.57	22.3	121	74
80	60	52	1.60	20.3	220	90
150	90	46	1.60	18.0	188	70
110	70	42	1.52	18.2	111	60
100	60	61	1.75	19.9	361	80
110	80	51	1.63	19.2	174	73
120	90	60	1.69	21.0	110	86
126	90	48	1.76	15.5	126	81
120	80	48	1.52	20.8	138	82
120	80	58	1.57	23.5	118	73
90	70	53	1.60	20.7	149	70
130	90	65	1.60	25.4	185	90
120	80	70	1.71	23.9	113	87
100	70	55	1.55	22.9	178	80
#NULL!	#NULL!	71	1.69	24.9	128	81
140	90	56	1.66	20.3	#NULL!	79
#NULL!	#NULL!	72	1.70	24.9	#NULL!	100
120	80	58	1.66	21.0	110	79
120	80	65	1.70	22.5	110	74
100	60	60	1.75	19.6	116	67
100	70	52	1.61	20.1	107	77
160	100	68	1.65	25.0	177	69
120	80	55	1.80	17.0	119	74
100	60	44	1.70	15.2	200	89
80	50	62	1.59	24.5	130	72

110	70	71	1.76	22.9	120	70
130	80	59	1.72	19.9	110	75
160	100	69	1.76	22.3	183	85
160	100	79	1.76	25.5	183	85
110	80	70	1.67	25.1	271	67
120	70	57	1.50	25.3	380	53
160	100	72	1.80	22.2	203	82
120	70	47	1.49	21.2	190	53
140	80	60	1.57	24.3	378	69
160	100	59	1.67	21.2	770	75
150	90	72	1.82	21.7	196	78
170	90	81	1.75	26.4	419	75
110	20	68	1.75	22.2	124	73
130	80	52	1.72	17.6	110	88
130	80	67	1.72	22.6	110	75
130	80	50	1.53	21.4	210	54
130	80	52	1.55	21.6	216	59
120	80	73	1.83	21.8	140	84
130	90	69	1.79	21.5	74	105
130	90	45	1.62	17.1	81	69
130	100	50	1.57	20.3	84	76
120	100	68	1.59	26.9	92	83
130	90	64	1.79	20.0	90	92
150	90	66	1.57	26.8	96	100
130	80	55	1.64	20.4	93	79
130	90	48	1.48	21.9	90	68
150	90	44	1.45	20.9	89	47
130	70	72	1.71	24.6	95	67
120	85	63	1.58	25.2	58	87
110	85	67	1.69	23.5	69	88
130	75	70	1.79	21.8	83	47
100	85	59	1.78	18.6	80	101
100	85	65	1.65	23.9	62	87
140	50	54	1.69	18.9	67	91
130	60	42	1.83	12.5	85	69
100	85	73	1.50	32.4	76	91
100	95	76	1.69	26.6	69	76
105	85	51	1.74	16.8	63	74
130	60	79	1.49	35.6	68	59
250	145	46	1.64	17.1	83	57
130	60	57	1.68	20.2	86	70
130	80	77	1.46	36.1	67	65
115	85	47	1.72	15.9	80	74
120	85	74	1.69	25.9	99	88

140	100	50	1.83	14.9	59	49
130	70	76	1.48	34.7	97	47
125	85	85	1.83	25.4	83	57
140	95	37	1.44	17.8	78	98
145	95	71	1.89	19.9	69	88
130	70	69	1.59	27.3	62	85
130	60	100	1.84	29.5	78	76
160	100	37	1.69	13.0	58	88
160	90	59	1.71	20.2	120	#NULL!
180	120	38	1.58	15.2	130	#NULL!
90	70	36	1.62	13.7	142	#NULL!
170	110	48	1.46	22.5	377	#NULL!
100	60	51	1.63	19.2	157	#NULL!
140	90	70	1.75	22.9	90	#NULL!
90	70	88	1.70	30.4	112	#NULL!
120	80	51	1.52	22.1	114	#NULL!
130	80	53	1.48	24.2	#NULL!	#NULL!
100	70	58	1.66	21.0	111	#NULL!
110	70	60	1.62	22.9	118	#NULL!
130	90	59	1.44	28.5	118	#NULL!
110	60	47	1.80	14.5	124	#NULL!
100	60	68	1.56	27.9	125	#NULL!
110	60	53	1.54	22.3	140	#NULL!
170	90	87	1.72	29.4	205	#NULL!
110	80	58	1.67	20.8	271	#NULL!
110	80	46	1.67	16.5	271	#NULL!
130	90	37	1.67	13.3	161	#NULL!
110	70	57	1.60	22.3	142	#NULL!
170	90	48	1.72	16.2	205	#NULL!
130	80	59	1.52	25.5	87	#NULL!
130	90	58	1.57	23.5	91	#NULL!
140	70	51	1.55	21.2	80	#NULL!
130	80	51	1.52	22.1	97	#NULL!
130	80	55	1.56	22.6	87	#NULL!
130	70	64	1.53	27.3	64	#NULL!
130	90	54	1.72	18.3	79	#NULL!
125	85	58	1.63	21.8	93	#NULL!
150	100	53	1.69	18.6	78	#NULL!
125	85	75	1.69	26.3	84	#NULL!
170	100	59	1.48	26.9	84	#NULL!
165	120	80	1.66	29.0	79	#NULL!
120	85	50	1.59	19.8	86	#NULL!
120	85	60	1.59	23.7	94	#NULL!
130	70	64	1.68	22.7	71	#NULL!

120	85	57	1.68	20.2	49	#NULL!
150	60	50	1.74	16.5	63	#NULL!
110	95	46	1.69	16.1	99	#NULL!
195	95	53	1.65	19.5	69	#NULL!
195	100	59	1.73	19.7	77	#NULL!
120	95	57	1.59	22.5	58	#NULL!
120	85	75	1.49	33.8	83	#NULL!
100	85	75	1.69	26.3	76	#NULL!
140	70	35	1.82	10.6	85	#NULL!
195	70	57	1.69	20.0	78	#NULL!
110	85	67	1.62	25.5	90	#NULL!
165	100	87	1.62	33.2	69	#NULL!
130	85	57	1.64	21.2	68	#NULL!
90	85	63	1.58	25.2	85	#NULL!
110	85	69	1.85	20.2	100	#NULL!
110	75	68	1.74	22.5	365	77
180	110	100	1.74	33.0	118	89
100	80	85	1.65	31.2	203	96
175	110	68	1.74	22.5	132	79
130	75	69	1.76	22.3	126	66
145	80	59	1.69	20.7	231	105
160	100	99	1.84	29.2	143	84
220	60	70	1.67	25.1	254	100
200	130	62	1.58	24.8	265	77
105	85	65	1.59	25.7	265	60
140	60	64	1.58	25.6	267	119
100	60	68	1.67	24.4	123	110
105	60	70	1.59	27.7	216	66
130	70	95	1.65	34.9	278	79
105	60	91	1.84	26.9	743	66
120	70	69	1.84	20.4	121	106
120	75	97	1.68	34.4	167	46
100	85	61	0.17	2110.7	114	109
220	100	70	1.54	29.5	117	78
120	50	63	1.60	24.6	154	121
125	65	64	1.74	21.1	324	106
120	60	66	1.69	23.1	342	95
120	75	73	1.64	27.1	467	106
140	100	68	1.84	20.1	321	82
125	60	68	1.69	23.8	115	75
130	75	70	1.85	20.5	88	94
130	85	66	1.79	20.6	73	50
125	60	68	1.67	24.4	69	66
110	75	68	1.48	31.0	113	55

130	60	67	1.85	19.6	129	106
105	75	64	1.78	20.2	119	77
120	60	68	1.69	23.8	94	86
120	85	57	1.75	18.6	79	87
120	65	60	1.73	20.0	234	79
125	75	65	1.68	23.0	58	83
105	60	54	1.82	16.3	100	81
105	60	57	1.73	19.0	59	109
120	75	57	1.64	21.2	100	79
100	75	76	1.73	25.4	324	110
195	95	60	1.63	22.6	576	66
115	60	57	1.74	18.8	96	79
145	85	98	1.73	32.7	123	88
100	60	60	1.72	20.3	217	86
120	85	65	1.57	26.4	227	86
120	60	57	1.64	21.2	208	79
130	75	92	1.73	30.7	217	69
135	70	55	1.68	19.5	117	76
125	85	75	1.63	28.2	118	76
105	60	94	1.68	33.3	398	89
125	75	54	1.68	19.1	889	78
120	85	56	1.75	18.3	341	66
105	75	50	1.67	17.9	278	76
115	70	57	1.63	21.5	321	75
125	85	52	1.68	18.4	117	87
120	85	86	1.74	28.4	119	108
125	56	60	1.68	21.3	287	107
100	57	54	1.76	17.4	321	86
110	60	94	1.83	28.1	334	79
110	85	50	1.83	14.9	324	108
100	60	65	1.58	26.0	219	78
120	60	48	1.68	17.0	118	85
115	75	65	1.62	24.8	1103	99
120	70	68	1.73	22.7	672	105
125	60	66	1.63	24.8	127	85
110	75	59	1.73	19.7	112	78
120	60	69	1.62	26.3	107	114
165	95	57	1.78	18.0	118	87
110	75	57	1.63	21.5	70	73
120	70	84	1.67	30.1	90	78
110	90	112	1.67	40.2	79	79
100	85	78	1.59	30.9	97	67
130	65	58	1.58	23.2	69	45
115	75	49	1.64	18.2	103	75

125	70	56	1.77	17.9	130	103
130	60	57	1.63	21.5	178	66
120	75	68	1.65	25.0	79	110
110	65	69	1.67	24.7	80	89
135	65	56	1.69	19.6	100	99
135	65	66	1.62	25.1	84	78
110	85	73	1.62	27.8	96	78
125	85	68	1.60	26.6	92	68
175	100	69	1.52	29.9	76	73
110	85	55	1.70	19.0	99	96
130	60	56	1.80	17.3	86	98
125	85	69	1.72	23.3	89	89
110	85	261	1.55	108.6	86	79
160	110	69	1.60	27.0	69	97
105	85	73	1.75	23.8	95	105
185	110	56	1.57	22.7	66	75
255	120	55	1.76	17.8	82	88
135	60	68	1.76	22.0	92	89
110	85	62	1.72	21.0	77	45
175	65	46	1.64	17.1	69	70
110	85	63	1.70	21.8	96	93
120	90	56	1.57	22.7	100	78
120	85	79	1.63	29.7	59	79
120	85	73	1.76	23.6	84	66
160	65	82	1.55	34.1	69	58
120	95	64	1.70	22.1	89	84
135	70	82	1.67	29.4	86	85
130	70	60	1.53	25.6	100	70
120	95	68	1.56	27.9	58	91
115	60	80	1.84	23.6	421	#NULL!
140	100	96	1.68	34.0	123	#NULL!
160	85	68	1.56	27.9	107	#NULL!
130	80	85	1.81	25.9	174	#NULL!
120	70	60	1.65	22.0	145	#NULL!
110	75	110	1.57	44.6	143	#NULL!
120	75	69	1.79	21.5	187	#NULL!
125	70	98	1.84	28.9	214	#NULL!
110	70	67	1.68	23.7	114	#NULL!
160	100	70	1.59	27.7	124	#NULL!
220	130	64	1.42	31.7	143	#NULL!
80	40	95	1.49	42.8	121	#NULL!
110	60	63	1.76	20.3	143	#NULL!
115	75	55	1.49	24.8	168	#NULL!
260	140	50	1.86	14.5	136	#NULL!

135	70	67	1.74	22.1	321	#NULL!
130	60	66	1.77	21.1	296	#NULL!
150	100	66	1.54	27.8	265	#NULL!
185	100	100	1.74	33.0	597	#NULL!
130	85	69	1.65	25.3	698	#NULL!
100	85	110	1.72	37.2	633	#NULL!
100	60	100	1.78	31.6	143	#NULL!
105	75	85	1.95	22.4	113	#NULL!
105	75	67	1.76	21.6	103	#NULL!
110	75	84	1.75	27.4	115	#NULL!
125	60	67	1.84	19.8	115	#NULL!
100	95	68	1.74	22.5	216	#NULL!
120	85	100	1.79	31.2	124	#NULL!
125	75	63	1.69	22.1	132	#NULL!
135	95	60	1.89	16.8	126	#NULL!
100	70	64	1.89	17.9	115	#NULL!
175	105	82	1.68	29.1	116	#NULL!
120	60	67	1.74	22.1	143	#NULL!
125	85	68	1.77	21.7	115	#NULL!
100	70	92	1.74	30.4	117	#NULL!
110	60	66	1.64	24.5	186	#NULL!
100	75	82	1.59	32.4	143	#NULL!
120	70	63	1.74	20.8	116	#NULL!
125	60	93	1.58	37.3	112	#NULL!
125	65	79	1.84	23.3	187	#NULL!
110	75	75	1.63	28.2	124	#NULL!
100	70	66	1.69	23.1	223	70
130	85	61	1.64	22.7	334	#NULL!
120	65	40	1.83	11.9	132	#NULL!
125	65	86	1.74	28.4	265	#NULL!
145	105	93	1.48	42.5	80	#NULL!
200	100	65	1.69	22.8	79	#NULL!
175	100	59	1.73	19.7	92	#NULL!
210	100	68	1.58	27.2	100	#NULL!
240	100	68	1.78	21.5	114	#NULL!
120	90	64	1.67	22.9	116	#NULL!
215	120	99	1.75	32.3	67	#NULL!
120	80	92	1.84	27.2	221	#NULL!
120	70	79	1.58	31.6	60	#NULL!
100	85	100	1.76	32.3	432	#NULL!
85	75	64	1.83	19.1	77	#NULL!
120	60	89	1.38	46.7	127	#NULL!
125	85	64	1.84	18.9	116	#NULL!
100	75	95	1.67	34.1	114	#NULL!

125	70	82	1.68	29.1	89	#NULL!
100	75	94	1.69	32.9	93	#NULL!
185	100	67	1.62	25.5	254	#NULL!
105	80	55	1.72	18.6	64	#NULL!
100	75	84	1.68	29.8	100	#NULL!
130	60	79	1.83	23.6	84	#NULL!
125	75	52	1.73	17.4	66	#NULL!
120	85	105	1.63	39.5	1100	#NULL!
100	85	58	1.68	20.5	213	#NULL!
155	100	83	1.59	32.8	123	#NULL!
145	100	95	1.62	36.2	113	#NULL!
135	85	63	1.68	22.3	127	#NULL!
145	95	99	1.74	32.7	376	#NULL!
120	70	54	1.68	19.1	286	#NULL!
125	70	102	1.64	37.9	278	#NULL!
95	60	56	1.67	20.1	325	#NULL!
125	95	66	1.63	24.8	201	#NULL!
120	65	77	1.69	27.0	223	#NULL!
100	85	56	1.65	20.6	983	#NULL!
135	75	59	1.73	19.7	890	#NULL!
115	60	87	1.73	29.1	845	#NULL!
100	85	53	1.57	21.5	378	#NULL!
125	80	91	1.64	33.8	325	#NULL!
100	60	69	71.71	0.0	252	#NULL!
110	60	95	1.82	28.7	388	#NULL!
130	75	69	1.63	26.0	234	#NULL!
125	70	57	1.72	19.3	235	#NULL!
120	95	80	1.62	30.5	326	#NULL!
110	70	50	1.58	20.0	345	#NULL!
115	60	60	1.74	19.8	332	#NULL!
120	75	50	1.64	18.6	254	#NULL!
115	90	110	1.62	41.9	243	#NULL!
100	60	77	1.68	27.3	234	#NULL!
110	75	50	1.59	19.8	333	#NULL!
120	65	83	1.64	30.9	221	#NULL!
125	70	66	1.72	22.3	128	#NULL!
120	60	78	1.59	30.9	128	#NULL!
120	65	68	1.73	22.7	127	#NULL!
215	110	91	1.63	34.3	129	#NULL!
100	65	65	1.68	23.0	127	#NULL!
175	110	68	1.73	22.7	182	#NULL!
110	60	69	1.59	27.3	329	#NULL!
115	70	64	1.74	21.1	442	#NULL!
130	70	69	1.73	23.1	275	#NULL!

110	60	75	1.67	26.9	126	#NULL!
120	70	74	1.59	29.3	184	#NULL!
110	70	67	1.63	25.2	792	#NULL!
100	75	75	1.68	26.6	119	#NULL!
145	100	65	1.72	22.0	96	#NULL!
120	70	50	1.72	16.9	85	#NULL!
125	80	54	1.62	20.6	80	#NULL!
135	100	53	1.72	17.9	89	#NULL!
100	80	57	1.74	18.8	95	#NULL!
100	75	54	1.67	19.4	79	#NULL!
185	110	60	1.63	22.6	79	#NULL!
110	65	46	1.73	15.4	78	#NULL!
110	85	68	1.72	23.0	100	#NULL!
110	60	75	1.68	26.6	130	#NULL!
195	110	86	1.68	30.5	96	#NULL!
120	56	68	1.78	21.5	85	#NULL!
100	60	68	1.83	20.3	90	#NULL!
100	60	114	1.67	40.9	96	#NULL!
130	70	69	1.55	28.7	99	#NULL!
130	65	71	1.35	39.0	93	#NULL!