KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF SCIENCE

DEPARTMENT OF MATHEMATICS

IMPACT OF MULTIMORBIDITY PATTERNS ON ELDERLY HOSPITALIZATION AND MORTALITY: A CASE STUDY OF KWADASO S.D.A HOSPITAL.

By

Tawiah Richard

BSc. Mathematical Science (Statistics Option)

A thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfilment for the degree of

MASTER OF PHILOSOPHY

May, 2013

DECLARATION

I hereby declare that this submission is my own work towards the MPhil and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

Tawiah Richard (PG6202611)	…….	…….....
Student's Name and ID	Signature	Date

Certified by:

Dr. FT. Oduro	…….	…….....
Supervisor	Signature	Date

Certified by:

Mr. F.K. Darkwah	…….	…….....
Head of Department	Signature	Date

DEDICATION

This thesis is dedicated to my mother Madam Janet Sarpong and my uncle Mr. Duodu Baffour.

ACKNOWLEDGEMENT

I would first and foremost thank the Almighty God for blessing, protecting and strengthening me through my period of study. May your name be exalted, honored and glorified.

I express my sincere gratitude to my supervisor Dr. F.T. Oduro for his time, continuous support, patience, motivation and immense knowledge towards the writing of this thesis. His matured experience and knowledge guided me in all the time of research and writing of this thesis.

Much gratitude also goes to my family Mr. Duodu Baffour, Madam Janet Sarpong, Mr. Nti Alexander, Mrs. Doris Frimpong and every one, for their indefatigable love and support throughout my life.

Obviously, I owe a debt of appreciation to Mr. Benjamin Boamah of the statistics department and Mavis Donkor of chronic education unit, both of Kwadaso S.D.A hospital, for providing data for this research work.

I am equally grateful to Mr. George Adjavon, Mr. Adugyamfi Maxwell, Benn Shadrack, Osei Dwumfour Charles, Linda Afia Nyarko, Sandra Okrah and all my friends and course mates who have assisted me in diverse ways. Asare Sledge, Yakubu Kanyiri, Adu Kofi Asamoah, Felix Baidoo Antwi and Boakye Sampson also deserve special thanks.

ABSTRACT

In this thesis, we aimed to identify multimorbidity patterns, study their impact on the length of hospitalization (LOH) until death and on mortality among elderly patients. The study utilized a sample of 984 elderly general clinic patients aged 50 years or older from data gathered from Kwadaso S.D.A hospital. The multimorbidity patterns were identified by exploratory tetrachoric factor analysis. Patients were assigned to any pattern if they had at least two diseases with factor loadings of 0.50 or more in absolute value on the corresponding pattern. However, to study the impact of the multimorbidity patterns, accelerated failure time (AFT) models with proportional hazards (PH) were used. The hazards were that of the exponential and Weibull models. Three multimorbidity patterns were identified: cardio-metabolic and pain disorders (CMPD), cardio-pulmonary disorders (CPD) and gastrointestinal, low back pain and anxiety disorders (GLAD). These patterns affected 52.2% of the entire sample. The Weibull model (AIC=239.7) provided a better fit when compared to the exponential model (AIC=259.3). Results from the Weibull model revealed that the median LOH until death was decreased by a factor of 0.46 and a factor of 0.45 for patients with CMPD and patients with CPD respectively when compared to patients with GLAD. Therefore, the estimated median LOH until death were 16 days for patients with CMPD and patients with CPD and 35 days for those with GLAD. Hospitalized elderly patients with these multimorbidity patterns, especially those with CPD and CMPD were vulnerable to increasing likelihood of mortality.

TABLE OF CONTENTS

CHAPTER 2

LITERATURE REVIEW

CHAPTER 3

METHODOLOGY

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFT	Accelerated Failure Time
AIC	Akaike's Information Criterion
CMPD	Cardio-Metabolic and Pain Disorders
CPD	Cardio-Pulmonary Disorders
GLAD	Gastrointestinal, Low Back Pain and Anxiety Disorders
LOH	Length of Hospitalization
MATLAB	Matrix Laboratory
MSA	Measure of Sampling Adequacy

OPD	Out Patient Department
PH	Proportional Hazards
SAS	Statistical Analysis System
SDA	Seventh Day Adventist
SPSS	Statistical Package for Social Scientists
VISTA	Visual Statistics

CHAPTER 1

INTRODUCTION

1.0 Introduction

This chapter gives the background of the study, problem statement, objectives, methodology, justification and organization of the study.

1.1 Background of the study

The co-existence of two or more chronic diseases in the same individual, a condition commonly known as multimorbidity (van den Akker, et al. 1996) has become a global phenomenon. Its prevalence has increased substantially in recent decades and would continue to increase in coming years in all countries of the world (Hartmann, et al. 2011). However, the reported prevalence of multimorbidity varies across studies and this variation is partly due to differences in the operational definition of multimorbidity, study populations (patients, general population), data collection methods, sources of data (surveys, administrative data), target age groups and diseases considered. For instance, a systematic review of various studies on the prevalence of multimorbidity in different countries published between 1980 and September 2010 revealed that the prevalence of multimorbidity varies from 3.5% to 98.5% in primary care and 13.1% to 71.8% in general population (Fortin, et al. 2012).

Multimorbidity rises with age, as a result it is common among elderly persons (Barnett, et al. 2012). In an Australian biomedical study, Taylor, et al. (2010), estimated the prevalence of multimorbidity as 4.4%, 15.0% and 39.2% in persons aged 20-39, 40-59

and 60 years or older respectively. Similarly, in a study by Prados-Torres, et al. (2012), the prevalence of multimorbidity was estimated as 13% in people aged 15-44 years, 43% in 45-64 and 67% in individuals aged 65 years or older in primary care in Spain. As a result, some studies on multimorbidity mainly focused on the elderly populations.

Chronic diseases that are common in elderly persons may occur jointly by chance. However, diseases that occur together frequently may result to clustering of major chronic diseases. In this regard, analysis and identification of disease clusters or multimorbidity patterns would help to discover which chronic diseases are associated with each other and which are also not in association.

Multimorbidity patterns have previously been identified statistically in some studies by using cluster analysis (John, et al. 2003; Cornell, et al. 2007). The main disadvantage with the approach of cluster analysis is that the association of diseases with multiple clusters or patterns is impossible, whereas in reality some diseases may be associated with more than one pattern. Therefore, Schäfer, et al. (2010), used exploratory tetrachoric factor analysis as a new approach of clustering diseases which allowed diseases to be associated with more than one pattern.

The impact of multimorbidity on health outcomes such as mortality, disability, quality of life, health care costs, health care utilization etc., have been studied in some studies using different statistical models. For instance, Hudon, et al. (2008), investigated the relationship between multimorbidity and physical activity levels, and long-term limitations on activity, self-rated general health, psychological distress, and physical activity levels in adults using multinomial regressions. Similarly, in a study by

Marengoni, et al. (2011), multivariate logistic regression models were used to assess the association of disability and diseases, in terms of multimorbidity and specific pairs of diseases among persons aged 75 years or over. Also, Hunger, et al. (2011), developed a generalized additive regression model to assess the independent effects and combined effects of six self-reported major chronic conditions on health related quality of life. However, from the context of multimorbidity and its related health outcomes, no literature exists on the use of survival analysis. Therefore, this study is the first to apply parametric survival models precisely, accelerated failure time models with proportional hazards to determine the impact of different multimorbidity patterns on the survival time of patients.

In Ghana, multimorbidity has been neglected in health systems. Health education, research, diseases preventive and control programs and clinical guidelines are predominantly based on single diseases. Therefore, literature on multimorbidity in Ghana is lacking. However, to commence and practically provide evidence of multimorbidity in elderly patients in Ghana, data for the present study was obtained from Kwadaso Seventh Day Adventist (S.D.A) hospital in Kumasi, the capital of Ashanti region of Ghana.

1.2 Problem Statement

Multimorbidity posses a wide range of challenges to population health worldwide. It accounts for poor physical function (Kadam, et al. 2007), impairments in quality of life (Hunger, et al. 2011), higher mortality (Gijsen, et al. 2011), increased medical cost and health care utilization (Glynn L.G, et al. 2011) among others. Among people aged 65 or

older this phenomenon is now a norm (Hartmann, et al. 2011). Agbosangaya, et al. (2012), indicated that old age, female sex, family structure (not living with children less than 16 years) and lower household income were independently associated with elevated odds of having multimorbidity. This study further reported that education was not a strong predictor of multimorbidity after adjusting for the aforementioned factors.

Despite the increasing prevalence of multimorbidity, clinical guidelines for managing the health of multimorbid patients in many countries are usually built around single diseases (Fortin, et al. 2005). This may lead to polypharmacy (Nobili, et al. 2011), which may consequently cause severe undesirable effects including adverse interaction between drugs and diseases. A study by Boyd, et al. (2005), found that applying single disease guidelines to a patient with five chronic conditions would result in the prescription of 19 doses of 12 different drugs, taken at five time points during the day, and carrying the risk of 10 attendant interactions or adverse events. According to Mangin, et al. (2012), drugs are among the top five causes of death in the United States hospitals and also account for about 17% of hospital admissions for people over 65 years of age.

People with multimorbidity are often seen in multiple sites of care, including emergency rooms, outpatient settings, specialty clinics, hospitals, nursing home and rehabilitation facilities, and assisted livings (Boyd, et al. 2010). Also, multimorbid patients are vulnerable to suboptimal quality care (Anderson G, Horvath, 2002) in that, a patient with multiple chronic diseases needs interdisciplinary care that coordinates the various treatments prescribed. However, coordination of care hardly occurs. Instead,

patients are often given conflicting directives from multiple health care providers and may be left to themselves to sort out the conflict.

The purpose of this thesis was to identify significantly occurring clusters of chronic diseases (i.e. multimorbidity patterns), study their impact on the length of hospitalization until death and analyze their impact on mortality among elderly patients.

1.3 Objectives of the study

i. To identify multimorbidity patterns among elderly patients.
ii. To study the impact of the multimorbidity patterns on the length of hospitalization until death among elderly patients.
iii. To analyze the impact of the multimorbidity patterns on mortality among elderly patients.

1.4 Methodology

The study utilized hospital administrative data which was gathered from Kwadaso S.D.A hospital. The data contained the demographic and diagnostic information of patients, who attended general clinic from October, 2011 to July, 2012. Patients aged 50 years or older with single or multiple chronic diseases formed the population of the study. Therefore, 984 subjects were sampled for this study using systematic sampling method.

A descriptive analysis of the sampled data was carried out by calculating frequencies of the demographic variables. The variable age was grouped into four categories: 50-54, 55-59, 60-64 and 65 years or older. In all 13 chronic diseases namely, arthritis, asthma,

gastritis, diabetes mellitus, kidney stones, congestive heart failure, peptic ulcer disease, anxiety, migraine, hypertension, stroke, low back pain, and liver cirrhosis/hepatoma were used in the study.

In order to identify multimorbidity patterns exploratory factor analysis was based on the tetrachoric correlations between artificially dichotomized diagnostic information of all the patients. Extraction of the factors was performed using the principal factor method. Oblique rotation of factor loadings matrix was utilized. Moreover, the number of factors retained was based on the eigenvalue greater than 1 rule and a disease was defined to be associated with a factor if it had a factor loading of at least 0.50 in absolute value. The resulting factors were interpreted as multimorbidity patterns (i.e. clusters of significantly associated chronic diseases). Prevalence of these patterns were calculated by assigning the patients to the patterns if they had at least two diseases with a factor loading of 0.50 or more in absolute value on the corresponding pattern. A chi-square test of homogeneity was used to test if the proportion of males and females were the same regarding the prevalence of the multimorbidity patterns. Again, this test was used to study the prevalence of the patterns over different age groups.

Moreover, the impact of the multimorbidity patterns on the length of hospitalization until death and mortality were analyzed using the accelerated failure time models with proportional hazards. The hazards were that of the exponential and Weibull models. The analyses of these models were based on a 5% level of statistical significance.

The tetrachoric correlations was performed in the ViSta 6.4 software package whilst exploratory factor analysis, exponential and Weibull models were analyzed using SAS 9.1. The chi-square test of homogeneity was performed in SPSS 16.0.

1.5 Justification

With an increasing ageing population and evidence that multimorbidity increases with age, combined with its adverse effects on health outcomes (mortality, disability, quality of life) and rising health care cost associated with existing clinical method, multimorbidity increasingly becomes an important issue in health care. In particular, the identification of multimorbidity patterns would enhance a better understanding of how some chronic disease occur together and also provide knowledge about the health care needs associated with different multimorbidity patterns. This would serve as guide to help improve the health of individuals with multimorbidity.

Additionally, studying the impact of different multimorbidity patterns would provide information as to whether one multimorbidity pattern is dangerous than the other. Such information could help make informed decisions in health systems in order to reduce multimorbid deaths. In summary, findings in this thesis would go a long way to help map out good strategies for the allocation of resources and the acquisition of the right and necessary medical equipments to enhance quality health care which would positively affect the life of individuals with multimorbidity.

1.6 Organization of the study

Chapter 1 is made up of the introduction, which comprises of the background of the study, problem statement, objectives of the study, methodology and justification of the study. Chapter 2 highlights on review of literature of ideas of different authors whose findings have been defined in relation to the topic under study. Chapter 3 focuses on methodological review in the light of statistical tools that are relevant to the analyses of the various data gathered. The statistical tools and models covered under this chapter includes, tetrachoric correlation coefficient, exploratory factor analysis, chi-square tests of homogeneity, accelerated failure time models with proportional hazards. Chapter 4 deals with data analysis and discussion of results, and Chapter 5 consists of conclusion and recommendations. The project report however ends with references and appendices in supportive to the researcher's investigation.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

In this section there is a review of the work of several authors regarding definition, concept of multimorbidity and various studies done to discover the prevalence, patterns and impact of multimorbidity particularly in the elderly population. Researches, empirical work and authors' opinion are looked at. Below are the focuses of the review.

- People with Multimorbidity: Prevalence
- Multimorbidity patterns and their prevalence
- Influence of socio-demographic and Socioeconomic variables on multimorbidity
- Impact of Multimorbidity

2.1 People with Multimorbidity: Prevalence

The prevalence of multimorbidity has often been investigated in few countries, particularly Australia (Britt, et al. 2008), Sweden (Marengoni, et al. 2008b), Canada (Fortin, et al. 2010) and others. However, available literature of multimorbidity for developing countries is limited.

A systematic review of various studies on the prevalence of multimorbidity in different countries published between 1980 and September 2010 revealed that the prevalence of multimorbidity varies from 3.5% to 98.5% in primary care and 13.1% to 71.8% in general population (Fortin, et al. 2012). The variation in the prevalence rates was partly due to the operational definition of multimorbidity, study populations (patients, samples

from general population), sources of data (e.g. surveys, administrative data, chart reviews), data collection methods, targeted age groups and diagnoses considered.

In Australia, Britt, et al. (2008), estimated the prevalence rate of multimorbidity in general practice patients and in the countries general population. This study used patients self reports and medical records. Chronic conditions were classified according to the Cumulative Illness Rating Scale morbidity domains. In the surveyed patients, the overall prevalence of multimorbidity was estimated as 37.1%, 29.0% of patients who attended a general practice and 25.5% of the general population. Fortin, et al. (2010) compared the prevalence rates of multimorbidity in primary care and in the general population of Canada. In the general population multimorbidity prevalence was estimated based on the co-occurrence of ≥2 and ≥3 diseases of the seven diseases listed in the general population survey, whilst in the primary care an open list of chronic diseases were used. The prevalence of multimorbidity was found to be higher in each age group in a primary care than in the general population. A study by van Oostrom, et al. (2012), presented an overview of the prevalence of multimorbidity and comorbidity of chronic diseases in the Dutch population using 7 years data (2002-2008) of 212,905 general practice patients. The overall prevalence of multimorbidity in the Dutch population was estimated as 13% and among those older than 55 years the prevalence was estimated as 37%. In a cross-sectional study in Germany of a sample of insured policy holders who were aged 65 years and over, about 62% of the sample had multimorbidity (van den Bussche, et al. 2011a). In this study multimorbidity was defined as the presence of at least 3 chronic diseases in an individual out of a list of 46 chronic diseases.

In the adult population, the prevalence of multimorbidity is reported to be high than individual chronic diseases in many studies. For example, Fortin, et al. (2005b), compared the prevalence of multimorbidity and the prevalence of three chronic diseases (asthma, hypertension and diabetes) using results from different published articles. The estimated prevalence of multimorbidity was 60% among individuals aged 55-74 years. This was much higher than that of asthma (6.5%), hypertension (29.6%) and diabetes (8.7%).

2.2 Multimorbidity Patterns and their prevalence

Different approaches have being used with regards to the identification of multimorbidity patterns. Some studies identified multimorbidity patterns based on grouping or pairing diseases that frequently occurred together. Amongst these studies includes Agborsangaya, et al. (2012), Britt, et al. (2008), and Marengoni, et al. (2011). Others analyzed triadic combinations of the most prevalent chronic diseases. For instance van den Bussche, et al. (2011), used triads of six most prevalent individual chronic diseases to ascertain the combination of diseases that are specific multimorbidity. These combinations consisting of hypertension, lipid metabolism disorders, chronic low back pain, diabetes mellitus, osteoarthritis and chronic ischemic heart disease covered the morbidity spectrum of 42% of the multimorbid individuals.

However, only few studies used cluster analysis in order to discover and explain the natural groupings or associations between chronic diseases in a particular population. In an effort to suggest a new approach to identify patterns of comorbidity and multimorbidity in American Indian elders aged 60 years and older living in rural

communities, cluster analysis was used by John, et al. (2003). The study results revealed that more than half of the respondents (57%) had 3 or more of 11 chronic diseases. Moreover, the cluster analysis method revealed four clusters of multimorbidity namely, cardiopulmonary, sensory-motor, depression, and arthritis. Similarly, Cluster analysis was used by Cornell, et al. (2007) as an illustrative approach to identify multimorbidity patterns in a set of 45 chronic illnesses in primary care patients. Six patterns namely metabolic, obesity, liver, neurovascular, stress and dual diagnosis pattern were identified. This study further indicated that cluster analysis appears to be a useful statistical technique for identifying multiple disease clusters and patterns of multimorbidity.

Some recent publications on multimorbidity employed exploratory factor analysis to explore statistically non-random patterns of multimorbidity in their study population. Almost all of these studies used tetrachoric correlation coefficient as measure of the joint occurrence of any two chronic diseases. In these studies the factors extracted were called multimorbidity patterns or multimorbidity clusters. Schäfer, et al. (2010), were the first who applied factor analysis as a statistical approach to explore patterns of multimorbidity in population of persons who were aged 65 years or older. Three patterns were identified: cardiovascular/metabolic disorders (prevalence female: 30%; male: 39%), anxiety/depression/ somatoform disorders and pain (34%; 22%), and neuropsychiatric disorders (6%; 0.8%). This study demonstrated the advantage of using factor analysis which allows diseases to be associated with more than one pattern, in contrast to cluster analysis. In a sample of persons aged 65-94 years with multimorbidity prevalence of 58.6%, Kirchberger, et al. (2012), used factor analysis to

explore multimorbidity patterns using self reported chronic health conditions in self administered questionnaire and standardized telephone interviews in Germany. Cardiovascular and metabolic disease, joint, liver, lung and eye diseases, mental and neurologic diseases, and gastrointestinal diseases and cancer were the identified multimorbidity patterns. Their study is the second study in which exploratory factor analysis was applied to identify multimorbidity patterns. Holden, et al. (2011), identified six patterns of multimorbidity among workers in 58 Australian-based companies using exploratory factor analysis. The study revealed that clusters do not fall neatly into organ or body systems. Similarly, Prados-Torres, et al. (2012), used factor analysis to identify five clinically important multimorbidity patterns. They were cardio-metabolic, psychiatric-substance abuse, mechanical-obesity-thyroidal, psychogeriatric and depressive patterns.

2.3 Influence of socio-demographic and socioeconomic variables on multimorbidity

Multimorbidity increases with age and is substantial among older adults (Barnett, et al. 2012). The prevalence of multimorbidity was assessed across three different age groups in a study by Taylor, et al. (2010). The study showed that the prevalence of multiple chronic conditions or multimorbidity in the age groups 20-39, 40-59 and 60 years or older were 4.4%, 15.0% and 39.2% respectively. Besides, it was observed that 42.1% of those with multimorbidity were less than 60 years of age. Similarly, in a study by Prados-Torres, et al. (2012), the prevalence of multimorbidity was estimated as 13% in people aged 15-44 years, 43% in 45-64 and 67% in those aged 65 years or older in 19 primary care centres in Spain. In addition Britt, et al. (2008), found that the prevalence of multimorbidity and complexity (number of domains present) increases with age:

83.2% of patients aged 75 years or over had multimorbidity, 58.2% had morbidity in three or more domains, and 33.4% in four or more. Arthritis/chronic back pain and vascular disease (15.0% of the sample), a psychological problem and vascular disease (10.6%), and arthritis/chronic back pain and a psychological problem (10.6%) were the common morbidity combinations.

Agborsangaya, et al. (2012), in their study estimated the prevalence multimorbidity and specific patterns of multimorbidity among adults aged 18 years or over in the general population of Alberta. Even though multimorbidity is known to increase with age, in this study the absolute number of individuals aged less than 65 years accounted for 70.2% of multimorbidity in the survey. However, this figure suggests that multimorbidity is not confined to the older adults.

Statistics about gender differences of the prevalence of multimorbidity is relative. Fortin, et al. (2005), in their study estimated the prevalence of multimorbidity across different age groups in a sample consisting of 320 males and 660 females of family practice patients in the Saguenay region (Quebec, Canada). The prevalence estimates were established by counting the number of chronic diseases and using the Cumulative Illness Rating Scale (CIRS) as a measure of severity for each of these diseases. The prevalence of two or more chronic diseases or multimorbidity in age groups 18-44, 45-64 and 65 years or older were estimated as 68%, 95%, and 99% among females and 72%, 89% and 97% among males. These figures suggest that there is no significant difference in the prevalence rate of multimorbidity across the age groups of both sexes. This study results on gender corresponds to that of Britt, et al. (2008). Conversely, Schäfer, et al. 2012, showed that the prevalence of multimorbidity appear to be high in

females than males across three age groups, thus 11% male and 16% female in age group 15-44, 39% male and 47% female in age group 45-64 and 65% male and 69% female in individuals aged 65 years or over. Similarly, Agborsangaya, et al. (2012), indicated in their work that the age-standardized prevalence of multimorbidity is higher in females (19.2%, 95% CI: 17.8-20.6) than in males (15.6%, 95% CI: 14.2-16.9).

In a prospective cohort study, Nagel, et al. (2008), investigated the statistical association between the levels of education attained and multimorbidity, by also taking into account intermediate factors that could explain such associations among individuals aged 50-75 years. The study used 13,781 individuals of the Heidelberg cohort of the European Prospective Investigation into Cancer and Nutrition. The overall prevalence of multimorbidity (having two or more chronic diseases) was 67.3%. Compared to the highest educational category, the lowest was found to be statistically significantly associated with increased odds of multimorbidity in men (OR= 1.43; 95% CI=1.28-1.61) and women (OR=1.33; 95% CI=1.18-1.57), however, increasing body mass index was identified as the most important predictor of this association. Additionally, poor education was found to be associated with increased risk of multimorbidity (Marengoni, 2008a). Generally, the impact of age, gender and socioeconomic status on the occurrence rate of chronic diseases and multimorbidity in 1099 elderly patients in Sweden was studied by Marengoni, et al. (2008b). The study showed that more than half (55%) of the participants had multimorbidity. However, advanced age, female gender, and lower level of education were independently associated with a more than 50% increased risk of multimorbidity. In Bangladesh the prevalence and distribution patterns of multimorbidity among elderly (≥60 years of age) rural dwellers was reported

in a study by Khanam, et al. (2011). Multimorbidity was defined as suffering from two or more of nine chronic medical conditions, such as arthritis, stroke, obesity, signs of thyroid hypofunction, obstructive pulmonary symptoms, symptoms of heart failure, impaired vision, hearing impairment and high blood pressure. The overall prevalence of multimorbidity in the study population was 53.8%, and it was reported to be higher among women, illiterates, persons who were single and persons in the non-poorest quintile. The results of multivariate logistic regression model developed in the study showed that female sex and belonging to non-poorest quintile were independently associated with an increased odds ratio of multimorbidity. Additionally, the association between socio-demographic factors and multimorbidity was examined in a study by Agborsangaya, et al. (2012). They found that multimorbidity was associated with sex, age, family structure and household income.

Schäfer, et al. (2012) developed a multilevel mixed-effect linear regression model to investigate the association between socio-demographic variables and especially socioeconomic status with general multimorbidity and with the specific patterns of multimorbidity among persons aged 65 years or older. Multimorbidity in general was found to be significantly associated with age (+0.07 chronic conditions per year), gender (-0.27 conditions for female), education (-0.26 conditions for medium and -0.29 conditions for high level vs. low level) and income (-0.27 conditions per logarithmic unit). With regards to specific multimorbidity patterns, cardiovascular/metabolic pattern was also associated with age (+0.04 chronic conditions per year). Variables including living arrangements and marital status were not associated with general multimorbidity. According to the study, two groups of elderly multimorbid patients were identified i.e.

firstly, those with mainly cardiovascular and metabolic disorders, who were often male, have an older age and a lower socioeconomic status and secondly those mainly with anxiety, depression, somatoform disorders and pain-related morbidity, who were more often female and equally distributed across age groups. Agborsangaya, et al. (2012), examined the statistical association between a range of socio-demographic factors and multimorbidity and found that age, sex, income and family structure were independently associated with multimorbidity.

Barnett, et al. 2012 examined the distribution of multimorbidity and comorbidity of physical and mental health disorders, in relation to age and socioeconomic deprivation using data extracted from a cross-sectional study in Scotland. In all, 40 diseases were considered and the analyses were based on the number of diseases, type of disorder (physical or mental), sex, age and socio-economic status. The study results showed that 42.2% (95% CI 42.2-42.3) of all patients had one or more diseases and 23.2% (95% CI 23.08-23.21) had two or more diseases or multimorbidity. Socio-economic deprivation was found to be associated with multimorbidity that included mental health disorders and was common among patients living in most deprived areas (11.0%) than those in least deprived areas (5.9%). A multivariate model developed by Taylor, et al. (2010), revealed that family structure, marital status, education level, country of birth, medication use, health service use, existence of depressive symptoms, smoking status, overall health status, high waist hip ratio and waist circumference are associated with multimorbidity.

In a zero inflated Poisson regression model developed by Tucker-Seeley, et al. (2011), childhood financial hardship and lifetime earnings were found to be associated with

multimorbidity, but not associated with the absence of morbidity in a cross sectional study of individuals who were aged 50 years or more in the United States. However, the association between childhood financial hardship and multimorbidity was seen to be influenced by lifetime earnings.

2.4 Impact of Multimorbidity

Multimorbidity impairs patients' quality of life, increases their risk of functional limitations and makes effective treatment difficult to deliver (Hartmann, et al. 2011). Multimorbidity is known to be inversely related to quality of life (QOL) or health related quality of life (HRQOL) and this relationship was confirmed in a study by Fortin, et al. (2004). Hunger, et al. (2011) developed a generalized additive regression model to assess the independent effects and combined effects of six self-reported major chronic conditions (thus, diabetes mellitus, coronary events, stroke, cancer, chronic bronchitis, hypertension) on health related quality of life in Germany. All the six diseases together with their interactions were inversely related to health related quality of life. The interaction effects (coefficient of interaction term -8.1, p <0.0001) of diabetes and coronary disorders in patients significantly resulted to more disability far greater than could be expected from their individual effects. Moreover, the combined effect of coronary disorders and stroke also showed a synergistic effect on health related quality of life.

Kadam, et al. (2007), in their study investigated the distribution of multimorbidity and its effects on the overall health of patients seen in family practice consultations over 18-month period. The target population of the study was those aged 50 years or more. The

distribution of multimorbidity was examined by applying two scales of multimorbidity, one based on simple morbidity counts and the other on the severity classification. They showed that 19% consulted for single morbidity and 23% for six or more (a high morbidity count) in the 18-month period. Besides, they found that high multimorbidity may account for about 24% of poor physical function in family practice consulting population. Again it was found that increasing severity of multimorbidity resulted to rising poor physical function. They concluded that multimorbidity defined by using routinely collected family practice consultation data and classified by count and by severity was associated with poorer physical function. Correspondingly, a study by Marengoni, (2008), showed that the number of chronic diseases incrementally increased the risk of functional decline, with Relative Risk (RR) increasing from 1.5 in individuals with one disease to 8.0 in persons with 5 or more diseases. Mortality occurred independently of the number of chronic conditions with Relative Risk 7.8 in subjects with one disease and 6.9 in those with multiple chronic disorders.

Hudon C, et al. (2008) investigated the relationship between multimorbidity and physical activity levels, and long-term limitations on activity, self-rated general health, psychological distress, and physical activity levels for each sex in adults, after age, education, income, and employment factors were controlled for. The study used a sample of 16,782 adults aged 18-69 years from data gathered from the Quebec Health Survey 1998. The associations between the dependent variable (physical activity levels) and the independent variables (multimorbidity, long-term limitations on activity, self-rated general health and psychological distress) were examined separately for males and females using multinomial regressions. The results revealed that about 46% of the

participants were males. Multimorbidity was not associated with physical activity levels for both males and females. Males and females with long-term limitations on activity and with poor-to-average self-rated general health were less likely to be physically active. Furthermore, there existed no statistical association between psychological distress and physical activity in males whereas in females, individuals with high levels of psychological distress are less likely to be physically active.

However, specific disease combinations or specific multimorbidity patterns are much more associated with disability than others. A study by Marengoni, et al. (2011) assessed the association of disability and diseases, in terms of multimorbidity and specific pairs of diseases among persons aged 75 years or over in Stockholm, Sweden. The study revealed that functional disability was present in 17.9% patients and was also found to increase with increasing number of chronic diseases. The prevalence of disability varied greatly amongst specific pairs of diseases. Disability was seen in 6.7% of individuals with hypertension and atrial fibrillation and 82.4% in persons with dementia and hip fracture. Analyses of multivariate logistic regression models revealed that dementia-hip fracture, dementia-cardiovascular disease, and dementia-depression were the pairs of diseases that significantly influenced the increased odds of functional disability among the aged population.

Multimorbidity inevitably leads to the use of multiple drugs, a condition known as polypharmacy (Nobili, et al. 2011). Vyas et al., (2012), estimated the rates of polypharmacy among multimorbid individuals aged above 21 years, having at least one physical condition in the following disease clusters: cardio-metabolic (diabetes or heart disease or hypertension), musculoskeletal (arthritis or osteoporosis), and respiratory

(chronic obstructive pulmonary disease (COPD) or asthma). Chi-square tests and logistic regression were used to analyze the association between multimorbidity and polypharmacy. The lowest polypharmacy rate (7.2%) was found in those with respiratory cluster. The rate of polypharmacy among those with all three clusters was 64.1%. However, the rate was higher among individuals with both cardio-metabolic and respiratory clusters (41.8%) than those with musculoskeletal and respiratory clusters. Individuals with cardio-metabolic conditions alone or in conjunction with other disease clusters were more likely to have polypharmacy. Compared to those with musculoskeletal and respiratory conditions, those with cardio-metabolic and respiratory and respiratory conditions had 1.68 times higher chance of polypharmacy. Polypharmacy can be appropriate, but it is associated with riskier prescribing and is often particularly problematic in persons who are physically frail or have cognitive impairments (Guthrie, et al. 2012). The influence of the number of chronic diseases in patients together with other two factors on the number of different prescriptions was studied by Laux, et al. (2008) using a multiple linear regression model. The model showed that the number of patients' chronic diseases significantly increases the number different prescriptions with parameter estimate 0.226. This suggests that a unit increase in the number of patients' chronic diseases increases the number of different prescriptions significantly by 22.6%.

People with multimorbidity have increased medical cost and health care utilization (Glynn, et al. 2011). In a sample of 3,309 primary care multimorbid patients who were aged 50 years or over in the west of Ireland, Glynn, et al. (2011), examined the prevalence of multimorbidity and associated health care utilization and cost. The

prevalence of multimorbidity was estimated as 66.2% (95% CI: 64.5-67.8). However, the presence of multimorbidity resulted to increased health care utilization and cost. Increasing number of chronic conditions resulted to significant increase in primary care consultations (p-value < 0.001, 11.9 versus 3.7 for >4 conditions versus 0 conditions), hospital out-patient visits (p-value < 0.001, 3.6 versus 0.6 for >4 conditions versus 0 conditions), hospital admissions (p-value < 0.001, adjusted odds ratio (OR) of 4.51 for >4 conditions versus 0 conditions) and total health care costs (p-value < 0.001, €4,096.86 versus €760.20 for >4 conditions versus 0 conditions). A study by van den Bussche, et al. (2011b), found that the presence of multimorbidity and nursing dependency due to disability were significantly associated with high levels of ambulatory medical care utilization in German. In this study multimorbidity was defined as the presence of 3 or more chronic conditions of a list of 46 most prevalent chronic conditions based on ICD 10 diagnosis. The statistical methodologies used were multidimensional frequency counts with standard deviations and confidence intervals, and multivariable linear regression. The results of the study showed that multimorbid patients had more than twice as many contacts per year with physicians than non-multimorbid patients (36 versus 16). About 5.7 different physicians were seen by multimorbid patients and 3.5 of them were seen by non-multimorbid patients per year. Moreover, they found that the number of contacts and of physician contacted increased gradually with the number of chronic conditions. The influence of gender and age of multimorbid patients on health care utilization was negligible.

On the other hand, in a multiple linear regression model developed by Laux, et al. (2008), the effect of age, gender and the number of chronic diseases on three response

variables (number of different prescriptions, number of referrals and number of encounters) used as a measure of health care utilization were investigated. The results obtained suggest that compared to gender, patients' age had strong influence on all the three indicators of health care utilization. Additionally, it was observed that the number of chronic diseases or multimorbidity was a significant positive predictor of the number of encounters with parameter estimate of 0.51. Again the effects of multimorbidity on the number of different prescriptions as well as the number of referrals were significantly estimated as 0.226 and 0.3 respectively.

CHAPTER 3

METHODOLOGY

3.0 Introduction

The present chapter focuses on the data, sampling technique, and how the study was conducted. Also, it shows the theoretical aspects of the statistical models and tests used in this study.

3.1 Description of data and variables

The study utilized administrative data, gathered from the Kwadaso S.D.A hospital in Kumasi. The data contained the demographic and diagnostic information of patients, who attended general clinic from October, 2011 to July, 2012. Some these patients were hospitalized due to the severity of their conditions. Patients aged 50 years or older with single or multiple chronic diseases formed the population for the study. A sample of 984 out of 3,032 patients was used in this study.

Demographic and diagnostic information were the needed variables. The demographic variables were the age and sex of patients, whilst the diagnostic variables were the chronic diseases obtained from patients consultation reports. Another variable used in this study was time: this variable was defined as the length of hospitalization (in days) until an event of interest death occurs.

3.2 Data Collection and Sampling Technique

Information obtained included diagnostic information, age and sex of the subjects. Also, for the subjects who were admitted into the hospital their length of hospital stay

until discharge or death were recorded. Generally, data collection was done in the following steps.

i. Data was first obtained from Microsoft Excel data files, which contained the diagnostic information of patients who were seen in general clinic consultation from October, 2011 to July, 2012. To acquire the needed information from the target population, diagnostic information of patients aged 50 years or older were extracted. In all diagnostic information of 3,032 subjects with varying medical conditions were extracted.

ii. The second step involved a sampling of 982 out of 3,032 subjects using systematic sampling technique. To obtain a sampling interval, the total number of subjects was divided by the sample size (3,032/984=3.08). As a result, the first subject was selected randomly between the first and the third subjects after which every third subject was selected until a total of 984 subjects were obtained. Each of these subjects was studied and all important data on the patients were carefully recorded. These included the sex, age and diagnostic information.

iii. To obtain the length of hospitalization until discharge or death for those who were admitted, the OPD numbers of the 984 patients were linked to the admissions and discharges register and all the necessary information were recorded.

3.3 Data Analysis

To facilitate the analysis of the data the variable age was grouped into four categories, i.e. 50-54, 55-59, 60-64 and 65 years or older. In all 13 chronic diseases were used: arthritis, asthma, gastritis, diabetes mellitus, kidney stones, congestive heart failure, peptic ulcer disease, anxiety, migraine, hypertension, stroke, low back pain, and liver cirrhosis/hepatoma.

To determine the pair-wise association between the abovementioned chronic diseases, tetrachoric correlation matrix was performed. In doing so, the diagnostic variable was coded in a binary format (0=nonexistence of a disease and 1=existence of a disease). However, dichotomous diagnoses were assumed to have underlying continuous latent characteristic. Tetrachoric correlation was used because it is the appropriate measure to handle correlation, when variables are dichotomous in nature (Kubinger, 2003).

Therefore, exploratory factor analysis was based on the matrix of tetrachoric correlations. Extraction of the factors was carried-out using the principal factor method. This method was chosen because it was assumed that factors would not explain the total variance in the diagnoses data. Additionally, it was assumed that factors would be associated, i.e. being in one multimorbidity pattern may influence the risk of being in another pattern as well, and hence oblique rotation of factor loading matrix was used. Moreover, the number of clusters or factors retained was based on the eigenvalue greater than 1 rule and a disease was defined to be associated with a factor if it had a factor loading of at least 0.50 in absolute value. The resulting factors were interpreted as multimorbidity patterns (i.e. clusters of significantly associated chronic diseases).

Prevalence of these patterns were computed by assigning patients to any of the patterns if they had at least two diseases with a factor loading of 0.50 or more in absolute values on the correspond pattern. Chi-Square test of homogeneity was applied to investigate if the prevalence the multimorbidity patterns were the same among males and females and also among the age groups 50-54, 55-59, 60-64 and 65 years or over.

However, it was assumed that the length of hospitalization until death of patients may be influenced by their respective multimorbidity patterns. Therefore, the impacts of the identified multimorbidity patterns on the length of hospitalization until death and on mortality were performed using accelerated failure time (AFT) models with proportional hazards. When performing the analysis the linear predictor was set equal to the intercept in the reference pattern (i.e. GLAD): this defines the baseline hazard. Because we were interest in the LOH until death, the LOH for subjects who were discharged were considered right censored (i.e. time become incomplete at the right side). In doing so the variable censor was given two codes, i.e. 0 and 1 to indicate whether LOH until death is known (non-censored) or not (right censored) respectively. The response variable was the log of the length of hospitalization in days until an event of interest, "death" occurs among hospitalized patients who had at least one of the identified multimorbidity patterns. The predictor variable was the patterns of multimorbidity with three categories namely, cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders. The analyses of these models were based on a 5% level of statistical significance.

The analyses of these statistical techniques were made possible by the use of the following statistical software packages. All descriptive statistics and chi-square tests

were executed with SPSS (version 16.0). Matrix of tetrachoric correlations was performed using ViSta (version 6.4) software package, and exploratory factor analysis and survival analysis were carried out with SAS (version 9.1). For the case of the exploratory factor analysis the matrix of tetrachoric correlations obtained from ViSta was used in SAS. All graphical procedures were carried out with MATLAB.

3.4 Tetrachoric Correlation

Tetrachoric correlation is a product-moment correlation between two unobserved quantitative variables that have been measured on a dichotomous scale (Pearson, 1900). Therefore, in this study the subjects diagnostic information were measured on a dichotomous scale, thus whether a chronic disease was diagnosed or not upon medical consultation. In doing so, diagnostic information was given binary codes, i.e. (0=nonexistence of disease and 1=existence of disease) for all the chronic diseases used in this study.

When a sample of N subjects has each been measured on two dichotomous variables the sampled data can be summarized in a 2×2 contingency table. Let a, b, c, d be the cell frequencies of the contingency table, where $(a + b)$ and $(c + d)$ represent the two row marginal frequency totals and $(a + c)$ and $(b + d)$ represent the two column marginal frequency totals (Brown, 1977).

Let $N = a + b + c + d$ the total table frequency. Let z_1 and z_2 denote the standard normal deviates corresponding to the marginal probabilities $(a + c)/N$ and $(a + b)/N$ respectively, that is,

$$z_1 = \Phi^{-1}\left(\frac{a+c}{N}\right) \tag{3.0}$$

and

$$z_2 = \Phi^{-1}\left(\frac{a+b}{N}\right) \tag{3.1}$$

where Φ is the cdf of the standard normal distribution. The tetrachoric correlation coefficient is the parameter value for which the volumes of the double dichotomous bivariate standard normal distribution equal to the joint probabilities of the contingency table. The joint probability is chosen to be the probability a/N, corresponding to the existence of both dichotomous variables. Then the tetrachoric correlation, r_t is the correlation that satisfies

$$\frac{a}{N} = \int_{-\infty}^{z_2}\int_{-\infty}^{z_1} \phi(X_1, X_2, r_t)\, dX_1\, dX_2 \tag{3.2}$$

where $\phi(X_1, X_2, r_t)$ is the bivariate normal $p.d.f.$ given by

$$\phi(X_1, X_2, r_t) = \left[2\pi(1-r_t^2)^{\frac{1}{2}}\right]^{-1} \exp\left[-\frac{X_1^2 - 2r_t X_1 X_2 + X_2^2}{2(1-r_t^2)}\right] \tag{3.3}$$

and where $x_1 = z_1$ and $x_2 = z_2$ define the line that divides the bivariate normal distribution into four quadrants with probabilities corresponding to the probabilities of the four cells in the 2×2 contingency table (Castellan, 1966). When only one cell has zero frequency, the zero is altered to 0.5 and all other cell frequencies are correspondingly adjusted by 0.5 to maintain the original row and column marginal frequency totals.

When $a = d = 0, r_t = -1$ and when $b = c = 0, r_t = +1$. When $z_1 = z_2 = 0$, then an explicit solution exists.

i.e. $$r_t = -cos\left(\frac{2\pi a}{N}\right) \tag{3.4}$$

In all other cases, r_t must be found by iteration as a root of Equation 3.0. Pearson (1900) and Everitt (1910) approximated the bivariate normal integral by the tetrachoric series expansion

$$I = \int_{-\infty}^{z_2}\int_{-\infty}^{z_1} \phi(x_1, x_2, r_t)\, dx_1\, dx_2$$

$$= \left(\frac{a+b}{N}\right)\left(\frac{a+c}{N}\right) + \sum_{j=1}^{\infty} \frac{r_t^j}{j!}\phi(z_1, z_2, 0) v_{j-1} w_{j-1}, \tag{3.5}$$

where $v_0 = 1, v_1 = z$, and $v_j = z_1 v_{j-1} - (j-1)v_{j-2}$ for $j > 1$, and $w_0 = 1$, $w_1 = z_2$, and $w_j = z_2 w_{j-1} - (j-1)w_{j-2}$ for $j > 1$, respectively.

The tetrachoric correlation coefficient is just the estimates of Pearson correlation coefficient between dichotomous variables (Juras, Pasarić, 2006).When the variables are many the matrix of tetrachoric correlations is computed between all the variables.

3.5 Factor Analysis

The essential purpose of factor analysis is to represent the variables $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_\mathrm{p}$ as linear combinations of few random variables $\mathrm{F}_1, \mathrm{F}_2, \cdots, \mathrm{F}_\mathrm{m}$ $(m < p)$ called factors. The factors are the underlying constructs or latent variables that "generate" the $\mathrm{X}'\mathrm{s}$. Like the original variables, factors vary from individual to individual; but unlike the variables, the factors cannot be measured or observed.

If the original variables $X_1, X_2, \cdots, X_p$ are at least moderately correlated, the basic dimensionality of the system is less than p. The goal factor analysis is to reduce the redundancy among the variables by using a smaller number of factors.

3.5.1 Orthogonal Factor Model

We assume a random sample $X_1, X_2, \cdots, X_p$ from a homogeneous population with mean vector μ and covariance matrix $\mathbf{\Sigma}$. The factor analysis model expresses each variable as a linear combination of the underlying common factors $F_1, F_2, \cdots, F_m$, with p additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$, called errors, or sometimes, specific factors. For $X_1, X_2, \cdots, X_p$ in any observation vector $\mathbf{X}$, the factor analysis model is as follows:

$$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2$$

$$\vdots$$

$$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p.$$

or, in matrix notation

$$\underset{(p\times 1)}{\mathbf{X} - \boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\ \underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}} \qquad (3.6)$$

where

$$\mathbf{X} = (X_1, X_2, \cdots, X_p)', \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)', \quad \mathbf{F} = (F_1, F_2, \cdots, F_m)',$$

$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)'$ and

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pm} \end{bmatrix}$$

The coefficient ℓ_{ij} is called the loading of the ith variable on the jth factor, so the matrix **L** is the matrix of factor loadings. Note that the ith specific factor ε_i is associated only with the ith response X_i. The p deviations $\mathrm{X}_1 - \mu_1,\ \mathrm{X}_2 - \mu_2, \ldots,\ \mathrm{X}_\mathrm{p} - \mu_\mathrm{p}$ are expressed in terms of $p + m$ random variables $\mathrm{F}_1, \mathrm{F}_2, \cdots, \mathrm{F}_\mathrm{m},\ \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\mathrm{p}$ which are unobservable. This distinguishes the factor model from the multiple regression model in which the independent variables whose position is occupied by **F** in Equation 3.6 can be observed.

With so many unobservable quantities, a direct verification of the factor model from observations on $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_\mathrm{p}$ is hopeless. However, with some additional assumptions about the random vectors **F** and $\boldsymbol{\varepsilon}$, the model in equation (3.6) implies certain covariance relationships, which can be checked.

We assume that

$$E(\mathbf{F}) = \underset{(m\times 1)}{\mathbf{0}} \tag{3.7}$$

$$\mathrm{Cov}(\mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \underset{(\mathrm{m}\times\mathrm{m})}{\mathbf{I}} \tag{3.8}$$

$$E(\boldsymbol{\varepsilon}) = \underset{(p\times 1)}{\mathbf{0}} \tag{3.9}$$

$$Cov(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \underset{(p\times p)}{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix} \tag{3.10}$$

and that $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent, so

$$Cov(\boldsymbol{\varepsilon}, \mathbf{F}) = E[\boldsymbol{\varepsilon}, \mathbf{F}'] = \underset{(p\times m)}{\mathbf{0}} \tag{3.11}$$

These assumptions and the relation in (3.6) constitute the orthogonal factor model. The orthogonal factor model implies a covariance structure for $\mathbf{X}$. From model (3.6),

$$(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = (\mathbf{LF}+\boldsymbol{\varepsilon})(\mathbf{LF}+\boldsymbol{\varepsilon})'$$

$$= (\mathbf{LF}+\boldsymbol{\varepsilon})((\mathbf{LF})'+\boldsymbol{\varepsilon}')$$

$$= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\varepsilon}'+\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$$

so that $$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'$$

$$= \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$

$$= \mathbf{LL}' + \boldsymbol{\Psi} \tag{3.12}$$

according to equation (3.10). Also by independence $\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E[\boldsymbol{\varepsilon}, \mathbf{F}'] = \mathbf{0}$. By the model in (3.6)

$$(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = (\mathbf{LF}+\boldsymbol{\varepsilon})\mathbf{F}' = \mathbf{LFF}' + \boldsymbol{\varepsilon}\mathbf{F}',$$

so $$\text{Cov}(\mathbf{X}, \mathbf{F}) = E(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = \mathbf{L}E(\mathbf{FF}') + E[\boldsymbol{\varepsilon}, \mathbf{F}'] = \mathbf{L}.$$

On the other hand, allowing the factors $\mathbf{F}$ to be correlated so that $\text{Cov}(\mathbf{F})$ is not diagonal gives the oblique factor model.

3.5.2 Covariance Structure for the Orthogonal Factor model

1. $\text{Cov}(\mathbf{X}) = \mathbf{LL}' + \boldsymbol{\Psi}$

or $\mathrm{Var}(\mathrm{X_i}) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i$

$$\mathrm{Cov}(\mathrm{X_i}, \mathrm{X_k}) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km} \qquad (3.13)$$

2. $\mathrm{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$

   or $\mathrm{Cov}(\mathrm{X_i}, \mathrm{F_j}) = \ell_{\mathrm{ij}}$ (3.14)

The model $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$ is linear in the common factors. If the p responses X are, in fact, related to the underlying factors, but the relationship is nonlinear, such as in

$\mathrm{X_1} - \mu_1 = \ell_{11}\mathrm{F_1}\mathrm{F_3} + \varepsilon_1,\ X_2 - \mu_2 = \ell_{21}F_2F_3 + \varepsilon_2,$ and so forth, then the covariance structure $\mathbf{LL}' + \boldsymbol{\Psi}$ may be inadequate**.**

The portion of variance of the ith variable contributed by the m common factors is called the ith communality. That portion of $\mathrm{Var}(\mathrm{X_i}) = \sigma_{ii}$ due to specific factor is often called the uniqueness, or specific variance. Denoting the ith communality by h_i^2, we see from equation (3.13) that

$$\underbrace{\sigma_{ii}}_{\mathrm{Var}(\mathrm{X_i})} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{communality} + \underbrace{\psi_i}_{specific\ variance}$$

or $$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 \cdots + \ell_{im}^2 \qquad (3.15)$$

and $$\sigma_{ii} = h_i^2 + \psi_i\,, \qquad i = 1,2,\cdots,p \qquad (3.16)$$

The ith communality is the sum of squares of the loadings of the ith variable on the m common factors.

3.5.3 Non-uniqueness of Factor loadings

The loadings in the model (3.6) can be multiplied by an orthogonal matrix without impairing their ability to reproduce the covariance matrix in $\mathbf{\Sigma = LL' + \Psi}$. To demonstrate this, let $\mathbf{T}$ be an arbitrary orthogonal matrix so that $\mathbf{TT' = I}$, and we insert $\mathbf{TT'}$ into the basic model (3.6) to obtain

$$\mathbf{X - \mu = LTT'F + \varepsilon}.$$

$$\mathbf{X - \mu = L^*F^* + \varepsilon},$$

where

$$\mathbf{L^* = LT}, \tag{3.17}$$

$$\mathbf{F^* = T'F} \tag{3.18}$$

If $\mathbf{L}$ in $\mathbf{\Sigma = LL' + \Psi}$ is replaced by $\mathbf{L^* = LT}$, we have

$$\mathbf{\Sigma = L^*L^{*\prime} + \Psi = LT(LT)' + \Psi}$$

$$\mathbf{= LTT'L' + \Psi = LL' + \Psi}, \tag{3.19}$$

since $\mathbf{TT' = I}$. Thus the new loadings $\mathbf{L^* = LT}$ in (3.17) reproduce the covariance matrix, just $\mathbf{L}$ does in (3.12):

$$\mathbf{\Sigma = L^*L^{*\prime} + \Psi = \ LL' + \Psi}.$$

The new factors $\mathbf{F^* = T'F}$ in (3.18) satisfy the assumptions (3.8), (3.9) and (3.11). The communalities are also unaffected by the transformation, $\mathbf{L^* = LT}$.

3.5.4 Methods of Estimation

The loadings of the factor model can be computed in diverse ways. The commonly known ones are; the principal component method, the principal factor method, the maximum likelihood method and the iterated principal factor method. Only the principal factor method was used in this study as a result it is the only estimation method that would be discussed.

3.5.5 Principal Factor Method

Given p dichotomous variables, we obtain the matrix of tetrachoric correlation, $\mathbf{R_t}$ which is just the estimate of the matrix of Pearson correlation between dichotomous variables. Therefore factoring the tetrachoric correlation matrix is the same as factoring the Pearson correlation matrix. The principal factor method uses an initial estimate $\widehat{\mathbf{\Psi}}$ and factors $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$ to obtain

$$\mathbf{R_t} - \widehat{\mathbf{\Psi}} \cong \hat{\mathbf{L}}\hat{\mathbf{L}}',$$

where $\hat{\mathbf{L}}$ is $p \times m$ and is obtained below by using eigenvalues eigenvectors of $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$.

$$\hat{\mathbf{L}} = \mathbf{C_1}\mathbf{D}_1^{1/2} = \left(\sqrt{\lambda_1}\mathbf{c_1}, \sqrt{\lambda_2}\mathbf{c_2}, \cdots, \sqrt{\lambda_m}\mathbf{c_m}\right) \qquad (3.20)$$

We define $\mathrm{D}_1 = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m)$ with m largest eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ and $\mathbf{C_1} = (\mathrm{c}_1, \mathrm{c}_2, \cdots, \mathrm{c}_m)$ containing the corresponding eigenvectors.

The diagonal elements of $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$ are the communalities, $\hat{h}_i^2 = 1 - \hat{\psi}_i$.

$$\mathbf{R_t} - \widehat{\mathbf{\Psi}} = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & \hat{h}_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & \hat{h}_p^2 \end{bmatrix} \tag{3.21}$$

A popular initial estimate for a communality in $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$ is $\hat{h}_i^2 = R_i^2$, the squared multiple correlation between X_i and the other $p - 1$ variables. This can be found as

$$\hat{h}_i^2 = R_i^2 = 1 - \frac{1}{r^{ii}}, \tag{3.22}$$

where r^{ii} is the ith diagonal element of $\mathbf{R_t}^{-1}$. To use equation (3.22) $\mathbf{R_t}$ must be nonsingular. If $\mathbf{R_t}$ is singular, we can use the absolute value or the square of the largest correlation in the ith row of $\mathbf{R_t}$ as an estimate of communality.

After obtaining communality estimates, we calculate eigenvalues and eigenvectors of $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$ and use (3.20) to obtain estimates of factor loadings, $\hat{\mathbf{L}}$. The proportion of variance explained by the jth factor is

$$\frac{\lambda_j}{\text{tr}(\mathbf{R_t} - \widehat{\mathbf{\Psi}})} = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i} \tag{3.23}$$

where λ_j is the jth eigenvalues of $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$. The matrix $\mathbf{R_t} - \widehat{\mathbf{\Psi}}$ is not necessarily positive semidefinite and will often have some small negative eigenvalues. In such a case the cumulative proportion of variance will exceed 1 and then decline to 1 as the negative eigenvalues are added.

3.5.6 Choosing the number of factors

Several criteria have been proposed for choosing the number factors (m). Four of these criteria are listed below.

1. Choose m equal to the number of factors necessary for the variance accounted for to achieve a predictive percentage, say 80%, of the total variance, $\mathbf{tr}(\mathbf{R})$.
2. Choose m equal to the number of eigenvalues greater than the average eigenvalue. For $\boldsymbol{R}$ the average is 1.
3. Use a scree test based on a plot of eigenvalues of $\boldsymbol{R}$. 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 lines begins.
4. Test the hypothesis that m is the correct number of factors, $\mathbf{H_0}: \boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$, where $\mathbf{L}$ is $p \times m$.

3.5.7 Factor Rotation

As demonstrated in section 3.5.3, the factor loadings (rows of $\mathbf{L}$) in the population model are unique only up to multiplication by an orthogonal matrix that rotates the loadings. The rotated loadings preserve the essential properties of the original loadings; they reproduce the covariance matrix and satisfy all basic assumptions.

If $\hat{\mathbf{L}}$ is the $p \times m$ matrix of estimated factor loadings obtained by any method (principal component, principal factor and so forth) then

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \text{ where } \mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I} \qquad (3.24)$$

is a $p \times m$ matrix of "rotated" loadings. Moreover, the estimated covariance matrix (or correlation matrix) remains unchanged, since

$$\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}} = \hat{\mathbf{L}}\mathbf{T}\mathbf{T}'\hat{\mathbf{L}} + \hat{\boldsymbol{\Psi}} = \hat{\mathbf{L}}^{*}\hat{\mathbf{L}}^{*\prime} + \hat{\boldsymbol{\Psi}} \qquad (3.25)$$

Equation (3.25) indicates that the residual matrix $\boldsymbol{S_n} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\boldsymbol{\Psi}} = \boldsymbol{S_n} - \hat{\mathbf{L}}^{*}\hat{\mathbf{L}}^{*\prime} - \widehat{\boldsymbol{\Psi}}$, remains unchanged. Moreover, the specific variances $\hat{\psi}_i$ and hence the communalities $\hat{h}_i^2$ are unaltered.

Geometrically, the loadings of the ith row of **L** constitute the coordinates of a point in the loading space corresponding to X_i. The goal of rotation is to place the axes close to as many points as possible. If there are clusters of points (corresponding to groupings of $X's$), we seek to move the axes in order to pass through or near these clusters. This would associate each group of variables with a factor (axis), and make interpretation more objective.

If a rotation in which every point is close to an axis, then each variable loads high on the factor corresponding to the axis and has small loadings on the remaining factors. Once this simple structure is established, we observe which variables are associated with each factor, and the factor is defined or named accordingly. The commonest types of factor rotation are orthogonal and oblique. However, for purpose of the contents of this thesis only the later would be considered.

3.5.8 Oblique Rotation

The term oblique rotation refers to a transformation in which the axes do not remain perpendicular. Instead of the orthogonal transformation matrix **T** used in (3.24), an oblique rotation uses a general nonsingular transformation matrix **Q** to obtain $\mathbf{Q} = \mathbf{Q}'\mathbf{F}$, and by $\mathbf{cov}(\mathbf{Ay}) = \mathbf{A\Sigma A}'$,

$$\mathbf{cov}(\mathbf{F}^*) = \mathbf{Q}'\mathbf{I}\mathbf{Q} = \mathbf{Q}'\mathbf{Q} \neq \mathbf{I} \qquad (3.26)$$

Thus the new factors are correlated. Since distances and angles are not preserved, the communalities for $\mathbf{F}^*$ are different from those for **F**.

Various analytical methods for achieving oblique rotations have been proposed and are available in program packages. Typically, the output of one of these procedures includes a pattern matrix, a structure matrix, and a matrix of correlations among the oblique factors. For interpretation, we would usually prefer the pattern matrix rather than the structure matrix. The loadings in the row of the pattern matrix are the natural coordinates of the point (variable) on the oblique axes and serve as coefficients in the model relating the variable to the factors.

3.5.9 Interpretation of Factors

It is learnt from the previous section that the usefulness of rotation is to aid interpretation. The goal is to achieve a simple structure in which each variable loads highly on only one factor, with small loadings on the other factors. In practice, this goal often fails, but rotation usually produces loadings that are closer to the desired simple structure.

Generally, identification and interpretation of factors is based on the magnitudes of the rotated loadings (in absolute value). However, to assess the significance of factor loadings, a threshold value of 0.3 has been advocated for many writers. For most successful applications, however, a critical value of 0.3 is too low and will result in variables of complexity greater than 1. A target of 0.5 or 0.6 is typically more useful.

3.5.10 Validity of the Factor Analysis Model

The commonest approach of checking the validity of the factor analysis model is to assess how close $\mathbf{R^{-1}}$ is to a diagonal matrix. To do so, Kaiser (1970) proposed a measure of sampling adequacy,

$$\mathbf{MSA} = \frac{\sum_{i \neq j} \boldsymbol{r_{ij}^2}}{\sum_{i \neq j} \boldsymbol{r_{ij}^2} + \sum_{i \neq j} \boldsymbol{q_{ij}^2}}, \qquad (3.27)$$

where r_{ij}^2 is the square of an element from R_t and q_{ij}^2 is the square of an element from

$\mathbf{Q} = \mathbf{DR^{-1}D}$, with $\mathbf{D} = \left[(\mathbf{diag}\,\mathbf{R^{-1}})^{1/2}\right]^{-1}$. As $\mathbf{R^{-1}}$ approaches a diagonal matrix, **MSA** approaches 1. Kaiser and Rice (1974) suggest that **MSA** should exceed 0.8 for satisfactory results to be expected. Generally, **MSA** below 0.5 suggest that R is unsuitable for factoring. In addition, the communality estimates after factoring should be fairly large.

3.6 Chi-square test of homogeneity

In the χ^2 test of homogeneity we test the claim that different populations have the same proportion of individuals with some characteristic. The test requires a contingency table and it is computed in a very similar fashion to the χ^2 test of independence. The test statistic for the χ^2 test of homogeneity is given by:

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad (3.28)$$

With $(m-1)(n-1)$ degrees of freedom, Where O_{ij} and E_{ij} are the cell observed and expected values respectively. The expected values are computed by:

$$E_{ij} = \frac{column_i total \times row_j total}{grandtotal}$$

3.7 Survival Analysis

Survival analysis is a useful statistical approach mostly used in medicine and engineering to describe the time until an event of interest, called 'death' or 'failure' occurs. Generally, in survival analysis the variable of interest is time until an event occur which could be years, months, weeks or days. Alternatively, time can refer to the age of an individual when an event occurs.

In survival analysis, the variable time is usually defined as survival time, because it gives the time that an individual has "survived" over some follow-up period. Analysis of survival data focuses on summarizing the main features of distribution, such as median or other quartiles of time to fail. Data on times until failure or survival times has two important features.

i. The times are non-negative and typically have skewed distributions with long tails.
ii. Some subjects may survive beyond the study period so that their actual failure times may not be known completely; in this case, and other cases where the failure times are not known completely, the data are said to be censored.

However, censored times can be right or left. When the survival time of a subject occurs within the study period, then survival time is considered uncensored. If a subject is

redrawn from the study survival time is right censored (i.e. time become incomplete at the right side). Subjects whose survival times commenced before the study is considered left censored. In the present study only right censored times were considered. However, in the coding scheme, codes 0 and 1 were used to indicate uncensored and right censored survival times respectively.

3.7.1 Survivor function

Let the continuous random variable T denote the survival time and let $f(t)$ denote its probability density function (p.d.f). Then the cumulative distribution function (cdf) $F(t) = \Pr(T \leq t)$, expresses the probability that the event has occurred by time t.

The complement of cumulative distribution function gives the survival function

$$s(t) = \Pr(T > t) = 1 - F(t) = \int_t^{\infty} f(x)dx. \qquad (3.29)$$

This gives the probability of being alive at time t, or more generally, the probability that the event of interest has not occurred by time t.

3.7.2 Hazard Function

An alternative characterization of the distribution of T is given by the hazard function, or instantaneous rate of occurrence of the event, defined as

$$h(t) = \frac{f(t)}{s(t)}. \qquad (3.30)$$

From equation (3.29), the $-f(t)$ is the derivative of $s(t)$. This suggests rewriting equation (3.30) as

$$h(t) = -\frac{d}{dt}\{\log[s(t)]\} \qquad (3.31)$$

Now integrating from 0 to t and introducing the boundary condition $s(0) = 1$ (since the event is sure not to have occurred by time 0). Equation (3.31) can be solved to obtain a formula for the probability of surviving to time t as a function of the hazard at all times up to t.

Hence $$s(t) = \exp[-H(t)] \qquad (3.32)$$

where

$$H(t) = \int_0^t h(x)dx$$

or $$H(t) = -\log[s(t)].$$

$H(t)$ is called the cumulative hazard function (cumulative risk) or integrated hazard function.

3.7.3 Proportional hazards models

Models of the form

$$h(t) = h_0(t)\exp(\beta X)$$

are called proportional hazard models and $h_0(t)$, which is the hazard function corresponding to the reference group for all the explanatory variables, is called the

baseline hazard. This reflects the underlying hazard for subjects with all covariates $X_1, \ldots, X_p$ equal to 0 (i.e. the "reference group"). The general form is:

$$h(t) = h_0(t)\exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right)$$

So when we substitute all of the $X_i's$ equal to zero, we get:

$$h(t) = h_0(t)\exp\left(\beta_1 \times 0 + \beta_2 \times 0 + \cdots + \beta_p \times 0\right)$$

$$= h_0(t)$$

3.7.4 Parametric Survival Models

Parametric survival models are class of models in which the distribution of the outcome (i.e. time to event) is specified in terms of unknown parameters. The survival time is assumed to follow a known distribution. The proportional hazards (PH) assumption in parametric survival models expresses the hazards in terms of a baseline hazard. Many parametric models are accelerated failure time (AFT) models in which survival time is modeled as a function of predictor variables.

3.7.5 Accelerated Failure Time (AFT) models

Let $\boldsymbol{T_i}$ be a random variable representing the survival time of the i-th subject. The general form of an accelerated failure time (AFT) model is:

$$\log(T_i) = \beta_{AFT} X_i + \sigma\epsilon \qquad (3.33)$$

where $\log(T_i)$ is the log of survival time, β_{AFT} is the vector of AFT model parameters corresponding to the covariate vector X_i, ϵ is a random "error" term, σ is a scale factor.

In general, the vector of covariates or explanatory variables may affect survival time and it may be continuous or discrete variable. It may also include possible interactions.

However, choosing different distributions for ϵ, we can obtain different parametric distributions such as Exponential, Weibull, Gamma, Log-logistic, Normal, Lognormal etc. The exponential and Weibull models accommodate both the PH and AFT assumptions.

3.7.6 Assumptions of parametric survival models

To assess the adequacy of a model it is necessary to check assumptions such as the distributional assumptions (e.g. exponential, Weibull, etc.) and the accelerated failure time assumptions. The following assumptions were considered in this study.

i. For the Exponential model, the negative log of the survival function $-\log[S(t))]$ is linear with survival time, (t) for all covariates.
ii. For the Weibull model, the log-log of the survival function $\log(-\log S(t))$ is linear with the log of survival time, $\log(t)$ for all covariates.
iii. The proportional hazards and accelerated failure time assumptions hold in both cases if the lines appear to have the same slope (i.e. are parallel).

3.7.7 Exponential Model

If failure times $T_i = T(X_i)$ follow an exponential distribution, then the resulting exponential model is

$$\log(T_i) = -\beta X_i + \epsilon \qquad (3.34)$$

Where $\boldsymbol{\epsilon}$ follows an extreme value distribution (which means that e^{ϵ} follows a unit exponential distribution).

The survivor function $s_i(t;\phi_i) = \exp(-\phi_i t)$ and ϕ_i is reparameterized as,

$\phi_i = \exp(\beta X_i)$.

Conversely, the hazard function is

$$h_i(t) = \phi_i$$

The hazard function does not depend on time t so the probability of failure in the time interval $[t, t+\delta t]$ is not related to how long the subject has already survived. This 'lack of memory' property may be a limitation because, in practice, the probability of failure often increases with time. In such situations, an accelerated failure time model, such as the Weibull distribution and others may be more appropriate.

The probability density function is given by

$$f(t) = \phi_i \exp(-\phi_i t) \tag{3.35}$$

3.7.8 Weibull model

The Weibull distribution has two parameters. The parameters λ and σ determine the shape of the distribution and the scale, respectively. If failure times $\boldsymbol{T_i} = \boldsymbol{T}(\boldsymbol{X_i})$ follow a Weibull distribution, then the resulting Weibull model is

$$\log(T_i) = -\sigma\beta X_i + \sigma\epsilon \tag{3.36}$$

where ϵ follows an extreme value distribution and $\sigma = {}^{1}/_{\lambda}$.

The survivor function is given by

$$s_i(t;\phi_i,\lambda) = \exp\left(-\phi_i t^{\lambda}\right) \tag{3.37}$$

where $\phi_i = \exp(\beta X_i)$,

also the hazard function is given by

$$h_i(t;\phi_i,\lambda) = \phi_i \lambda t^{\lambda-1} \tag{3.38}$$

and the probability density function is

$$f(t) = \phi_i \lambda t^{\lambda-1} \exp\left(-\phi_i t^{\lambda}\right). \tag{3.39}$$

The hazard function depends on time t and with suitable values of λ it can increase or decrease. Thus, the Weibull distribution yields accelerated failure time model. If the shape parameter, $\lambda = 1$, then the hazard function remain constant over time and the Weibull model reduces to an exponential model. If $\lambda > 1$, then hazard increases over time and when $\lambda < 1$ the hazard decreases as time increases.

3.7.9 Estimating Survival time

The parameter estimates obtained from an exponential, Weibull or any parametric model can be used to estimate the time $\hat{t}$ to any value $s(t) = q$.

$$\hat{t} = [-\ln(q)] \times \exp(\beta_0 + \beta_1 X_1) \tag{3.40}$$

Where q, represents either the first, second (median) and third quartile of the data.

3.7.10 Choosing the appropriate parametric survival model

Choosing the most appropriate parametric model involves checking and validating the assumptions of the model which is discussed graphically in section 3.3.20. Akaike's information criterion (AIC) provides an approach for comparing models with different underlying distributions, making use of $-2 * \log \text{likelihood}$ statistic. The likelihood for any parametric model is a function of the observed data and the model's unknown parameters. The form of the likelihood is based on the probability density function $f(t)$ of the response variable. The AIC statistic is calculated as

$$AIC = -2 * \log \text{likelihood} + 2p \qquad (3.41)$$

(where p is the number of parameters in the model). A smaller AIC statistic suggests a better fit. The addition of 2 times p can be thought of as a penalty if non-predictive parameters are added to the model.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.0 Introduction

In this chapter, the analysis and results obtained by using the various statistical tools and procedures described in the previous chapter were presented. It includes a brief descriptive analysis of the demographic characteristics of the subjects and the prevalence of different chronic diseases among the subjects. The main results were achieved by exploratory tetrachoric factor analysis, exponential and Weibull models. The exponential and the Weibull models were based on a 5% level of statistical significance.

4.1 Descriptive Analysis

Table 4.1: Demographic characteristics of the sample (N=984)

	Age groups									
	50-54 years		55-59 years		60-64 years		≥ 65 years		Overall	
Sex	n	%	n	%	n	%	n	%	n	%
Males	83	8.4	126	12.8	82	8.3	125	12.7	416	42.3
Females	172	17.5	85	8.6	156	15.9	155	15.8	568	57.7

The demographic characteristics of the study sample are shown in Table 4.1. Of the 984 patients who were aged 50 years or older, 42.3% were male. Among them their modal age group was 55-59 years. On the other hand, 57.7% were female and their modal age group was 50-54 years.

Table 4.2: Prevalence the of 13 chronic diseases among the elderly patients

Chronic conditions	Prevalence (%)
Arthritis	15.2
Asthma	9.6
Gastritis	11.3
Diabetes mellitus	36.3
Kidney stones	1.7
Congestive heart failure	28.2
Peptic ulcer disease	18.5
Anxiety	7.2
Migraine	25.8
Hypertension	51.6
Stroke	20.4
Low back pain	10.5
Liver cirrhosis/ Hepatoma	1.3

Table 4.2 shows the prevalence of 13 chronic diseases among the elderly patients. Hypertension, diabetes mellitus, congestive heart failure, migraine, stroke, peptic ulcer disease, arthritis, gastritis, and low back pain were the most common diseases with prevalence 51.6%, 36.3%, 28.2%, 25.8%, 20.4%, 18.5%, 15.2%, 11.3%, and 10.5% respectively. The prevalence of asthma, anxiety, kidney stones and liver cirrhosis/hepatoma were less than 10%. Among these patients, about 31.9%, 30.0%, 25.3%, 8.9% and 3.9% were living with one, two, three, four and five or more chronic

diseases respectively (Appendix). The overall prevalence of multimorbidity was 68.1% in the entire sample.

The tetrachoric correlations matrix shown in table A3 in the appendix depicts the pair-wise associations between different chronic diseases. It clearly shows that some diseases were strongly associated whilst others are not. The strongest associations identified were diabetes mellitus and hypertension (0.83), asthma and congestive heart failure (0.81), gastritis and low back pain (0.81), stroke and migraine (0.80). Others including stroke and asthma (0.75), migraine and hypertension (-0.60), diabetes mellitus and migraine (-0.55), low back pain and kidney stones (-0.50), gastritis and kidney stones (0.47), gastritis and migraine (-0.46), diabetes mellitus and kidney stones (-0.43), etc were moderately correlated. Some associations identified were also weak. These include arthritis and congestive heart failure (0.03), liver cirrhosis/hepatoma and congestive heart failure (0.04), gastritis and congestive heart failure (-0.08), arthritis and gastritis (-0.12) among others.

4.2 Exploratory factor analysis based on tetrachoric correlations matrix

Table 4.3: Number of factors retained by eigenvalue greater than 1 rule

	Eigenvalue	Differences	Proportion	Cumulative
1	**4.3853**	2.4049	0.4751	0.4751
2	**1.9803**	0.7195	0.2146	0.6897
3	**1.2608**	0.2777	0.1366	**0.8263**
4	0.9831	0.4921	0.1065	0.9328
5	0.4910	0.2916	0.0532	0.9860
6	0.1994	0.0396	0.0216	1.0076
7	0.1598	0.0261	0.0173	1.0249
8	0.1337	0.0859	0.0145	1.0394
9	0.0479	0.0902	0.0052	1.0446
10	-0.0423	0.0260	-0.0046	1.0400
11	-0.0683	0.0391	-0.0074	1.0326
12	-0.1073	0.0863	-0.0116	1.0210
13	-0.1936		-0.0210	1.0000

Kaiser's Measure of Sampling Adequacy: Overall = 0.57

Table 4.3 provides the details of the number of factor retained in the factor analysis model. The first three eigenvalues 4.3853, 1.9803 and 1.2608 were the only eigenvalues greater than 1. Therefore, three factors were retained on the basis of eigenvalues greater than 1 rule. These factors accounted for a cumulative proportion of 0.8263 of total

variance. The overall Kaiser's Measure of Sampling Adequacy was 0.57. These figures suggest that factor analysis model is reasonable.

Table 4.4: Oblique Rotated Loadings of the factors retained

Chronic diseases	Factor 1	Factor 2	Factor 3
Arthritis	**-0.5345**	0.1828	0.1042
Asthma	-0.1269	**0.8960**	-0.0858
Liver cirrhosis/Hepatoma	-0.2266	0.0941	-0.2967
Diabetes mellitus	**0.8944**	0.1684	-0.0198
Gastritis	0.0427	-0.0545	**0.7458**
Congestive heart failure	0.3319	**0.9805**	0.0152
Peptic ulcer disease	-0.0901	0.3239	**0.5081**
Anxiety	0.1332	0.0388	**-0.5266**
Migraine	**-0.5712**	0.1521	-0.4656
Hypertension	**0.8943**	0.1811	0.0440
Stroke	-0.4001	**0.5516**	-0.3097
Low back pain	0.1779	-0.0898	**0.6924**
Kidney stones	-0.2850	0.1854	-0.4212

Table 4.4 shows the factor loadings of all the 13 chronic diseases after application of oblique rotation. The first factor is characterized by high loadings for diabetes mellitus and hypertension and moderate loadings for arthritis and migraine, therefore this could be interpreted as cardio-metabolic and pain disorders. The second factor could be named cardio-pulmonary disorders due to the substantial loadings for asthma,

congestive heart failure, and moderate loading for stroke. Finally, the third factor is recognized by moderate loadings for gastritis, peptic ulcer disease, anxiety and low back pain, therefore named as gastrointestinal, low back pain and anxiety disorders. Liver cirrhosis/hepatoma and kidney stones were not associated with any of the factors. Basically, three multimorbidity patterns were identified in the distribution of 13 chronic diseases among the elderly patients. These were;

- Factor 1: cardio-metabolic and pain disorders (CMPD)
- Factor 2: cardio-pulmonary disorders (CPD)
- Factor 3: gastrointestinal, low back pain and anxiety disorders (GLAD)

Table 4.5: Prevalence of the multimorbidity patterns stratified by sex and age

	CMPD	**CPD**	**GLAD**	**p-value**
Sex				0.005
Male (23.0%)	14.2%	5.3%	3.5%	
Female (29.2%)	13.9%	8.6%	6.6%	
Age				0.176
50-54	2.7%	2.1%	1.3%	
55-59	2.8%	1.9%	1.7%	
60-64	8.9%	4.5%	2.6%	
≥65	13.6%	5.3%	4.4%	
Overall (52.2%)	28.2%	13.9%	10.1%	

Table 4.5 depicts the prevalence of cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders and also

shows a chi-square test of homogeneity for the presence of these multimorbidity patterns among males and females and across the age groups 50-54, 55-59, 60-64 and 65 years or older. Cardio-metabolic and pain disorders was the most common pattern, with total prevalence of 28.2%, followed by cardio-pulmonary disorders (13.9%) and lastly by gastrointestinal, low back pain and anxiety disorders also with total prevalence of 10.1%. Generally, 52.2% of the elderly general clinic patients had at least one of the three patterns of multimorbidity. However, about 15.9% of the multimorbid patients could not be assigned to any of the patterns of multimorbidity.

The prevalence of cardio-metabolic and pain disorders was 14.2% in males and 13.9% in females. Also the prevalence of cardio-pulmonary disorders was 5.3% in males and 8.6% in females and for gastrointestinal, low back pain and anxiety disorders, its prevalence rates were estimated as 3.5% in males and 6.6% in females. The p-value of the chi-square statistic of 0.005 implies that the test is significantly different from zero at 5% level of statistical significance. This leads to the rejection of the null hypothesis in favor of the alternative hypothesis, which brings out the implication that the prevalence rates of the identified multimorbidity patterns are not the same among males and females. However, our results show that the multimorbidity patterns were common among females (29.2%) than males (23.0%).

The prevalence of cardio-metabolic and pain disorders was 2.7%, 2.8%, 8.9% and 13.6% in patients aged 50-54, 55-59, 60-64 and 65 years or older. In the same age groups the prevalence of cardio-pulmonary disorders was estimated respectively as 2.1%, 1.9%, 4.5% and 5.3%. Prevalence rates were 1.3%, 1.7%, 2.6% and 4.4% for gastrointestinal, low back pain and anxiety disorders among patients aged 50-54,

55-59, 60-64 and 65 years or older. However, the p-value of the chi-square statistic of 0.176 indicates that the test is not significant at 5% level. However, this suggests that the proportions of people with the multimorbidity patterns are the same across the age groups under consideration.

4.3 Association of the multimorbidity patterns and the log of LOH until death

In this section of the chapter, the AFT forms of the exponential and Weibull models were used to assess the statistical association between the multimorbidity patterns and the log of the length of hospitalization until death. The aim was to determine the quantitative effect of different multimorbidity patterns on the length of hospitalization until death and analyze their impact on mortality among elderly patients.

Table 4.6: Analysis of Exponential Model Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	4.5827	0.3333	3.9294	5.2360	189.01	< **0.0001**
CMPD	1	-1.2811	0.3773	-2.0206	-0.5416	11.53	**0.0007**
CPD	1	-1.3304	0.4014	-2.1171	-0.5437	10.99	**0.0009**
GLAD	0	0.0000	.	.	.	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000		
Weibull Shape	0	1.0000	0.0000	1.0000	1.0000		

Table 4.6 shows the results of the exponential model. It contains the parameter estimates for the multimorbidity patterns, their standard errors, confidence limits, p-values etc. The parameter estimate of the intercept was 4.5827. This is the baseline hazard and it reflects the underlying hazard for individuals with gastrointestinal, low back pain and anxiety disorders. The parameter estimates for cardio-metabolic and pain disorders, and cardio-pulmonary disorders were -1.2811 and -1.3304 respectively. The p-values indicate that all the three multimorbidity patterns are significantly different from zero at 5% level of statistical significance. Therefore, this leads to the conclusion that the multimorbidity patterns are significantly associated with the log of the length of hospitalization until death.

4.3.1 Exponential Acceleration Factor for the multimorbidity patterns

The results obtained below shows that, the acceleration factor for patients with cardio-metabolic and pain disorders compared to those with gastrointestinal, low back pain and anxiety disorders was 0.28, and that of patients with cardio-pulmonary disorders compared to those with gastrointestinal, low back pain and anxiety disorders was 0.26.

$$\hat{\gamma}(CMPD\ Vs.\ GLAD) = \exp\{-1.2811\} = 0.28$$

$$\hat{\gamma}(CPD\ Vs.\ GLAD) = \exp\{-1.3304\} = 0.26$$

These results suggest that the median (or any other quartile of the) length of hospitalization was decreased by a factor of 0.28 for patients with cardio-metabolic and pain disorders compared to those with gastrointestinal, low back pain and anxiety disorders. Also, among patients with cardio-pulmonary disorders compared to those

with gastrointestinal, low back pain and anxiety disorders the median (or any other quartile of the) length of hospitalization was decreased by a factor of 0.26.

4.3.2 Exponential Survivor functions for the multimorbidity patterns

The probability of surviving longer than t days for hospitalized patients with cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal low back pain and anxiety disorders can be calculated by the following survivor functions respectively:

$$s(t, CMPD) = exp\{-(0.0368 * t)^{1.0000}\}$$

$$s(t, CPD) = exp\{-(0.0387 * t)^{1.0000}\}$$

$$s(t, GLAD) = exp\{-(0.0102 * t)^{1.000}\}$$

The estimated survivor function of the exponential model is shown graphically in Figure 4.1. The graph depicts that the survival probabilities of patients with the patterns of multimorbidity decreases over time. Generally, individuals who suffered from gastrointestinal, low back pain and anxiety disorders have consistently higher survival probabilities than those with cardio-metabolic and pain disorders, and cardio-pulmonary disorders. In other words, individuals with gastrointestinal, low back pain, anxiety disorders have longer survival times as compared to persons admitted with cardio-metabolic and pain disorders, and cardio-pulmonary disorders.

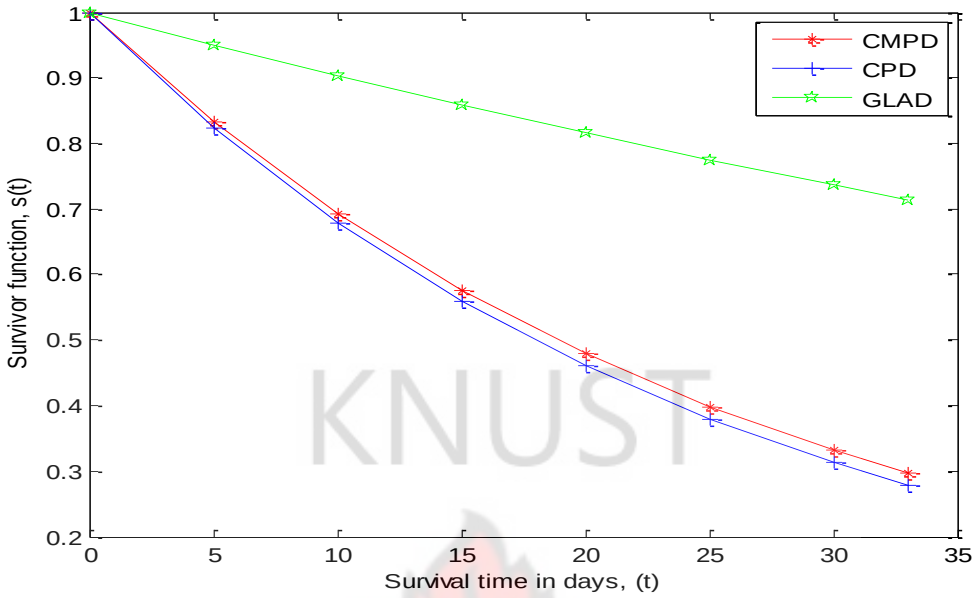

Figure 4.1: Exponential Survivor curves for the multimorbidity patterns

Table 4.7: LOH until death estimated from the exponential model

Quartile	S(t)	CMPD	CPD	GLAD
First quartile	0.25	37.6	35.8	135.5
Second quartile (Median)	0.50	18.8	17.9	67.8
Third quartile	0.75	7.8	6.4	28.1

Table 4.7 displays the estimated length of hospitalization for the persons who suffered from the multimorbidity patterns. The median length of hospitalization were 19 days, 18 days and 68 days for patients with cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders.

4.3.3 Exponential hazard functions for the multimorbidity patterns

From Table 4.6 the Weibull shape parameter of the exponential model is 1.00. This suggests that the hazard rate remains constant over time for all the multimorbidity patterns. This is demonstrated below and also shown graphically in figure 4.2.

$$h(t, CMPD) = 1.0 * 0.0368 * t^{1.0-1.0} = 0.0368$$

$$h(t, CMPD) = 1.0 * 0.0387 * t^{1.0-1.0} = 0.0387$$

$$h(t, GLAD) = 1.0 * 0.0102 * t^{1.0-1.0} = 0.0102.$$

The hazard probabilities for patients with cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders were 0.037, 0.039 and 0.010 respectively and these figures remain constant over time. These suggest that the likelihood of mortality was independent of the length of hospitalization.

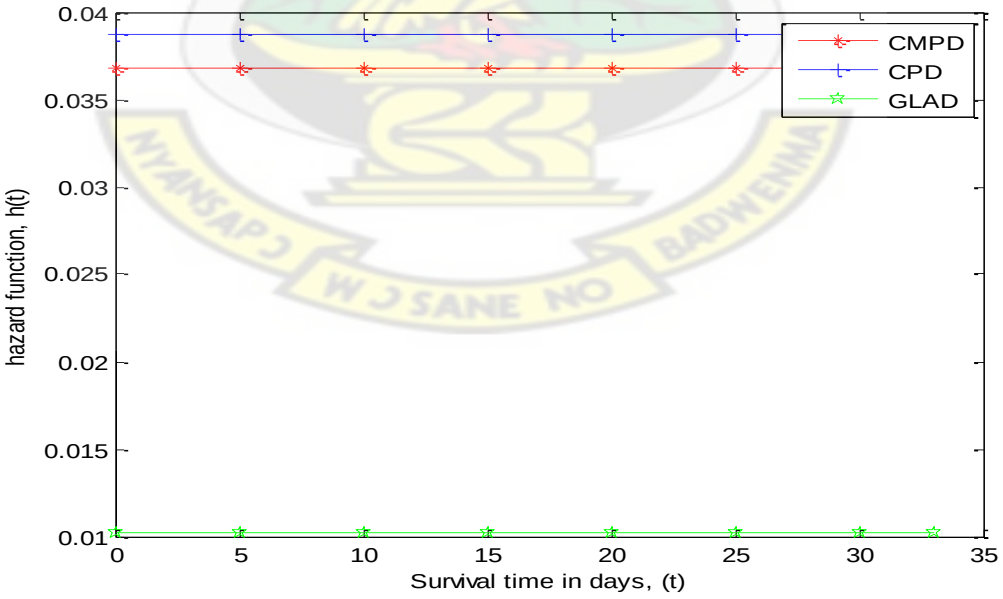

Figure 4.2: Exponential hazard curves for the multimorbidity patterns

Figure A1 in the appendix displays the plot of the negative log of the estimated survivor functions, against the survival time. The resulting plots are straight for all the three multimorbidity patterns, therefore, suggesting that the exponential model is reasonable. Furthermore, the lines appear to have different slopes (i.e. not parallel) suggesting that both PH and AFT assumptions are violated. Therefore, the exponential model is inappropriate even though the survival data follows the exponential distribution.

Table 4.8: Analysis of Weibull Model Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.9246	0.2189	3.4956	4.3537	321.48	**< 0.0001**
CMPD	1	-0.7739	0.2363	-1.2371	-0.3107	10.72	**0.0011**
CPD	1	-0.7992	0.2513	-1.2916	-0.3067	10.12	**0.0015**
GLAD	0	0.0000	.	.	.	.	.
Scale	1	0.5903	0.0634	0.4782	0.7287		
Weibull Shape	1	1.6940	0.1820	1.3724	2.0911		

Table 4.8 shows the results of Weibull model. It contains the Weibull parameter estimates, their standard errors, confidence limits, Weibull shape parameter, p-values etc. The parameter estimates for the intercept, cardio-metabolic and pain disorders, and cardio-pulmonary disorders were 3.9246, -0.7739 and -0.7992 respectively. However, the parameter estimate for the intercept is the baseline hazard and it reflects the underlying hazard for individuals with gastrointestinal, low back pain and anxiety

disorders. The intercept together with the slope parameters for cardio-metabolic and pain disorders, and cardio-pulmonary disorders are all significantly associated with the log of length of hospitalization until death at 5% level of statistical significance.

4.3.4 Weibull Acceleration Factor for the multimorbidity patterns

The results obtained below shows that the acceleration factor for cardio-metabolic and pain disorders compared to gastrointestinal, low back pain and anxiety disorders was 0.46, and that of cardio-pulmonary disorders compared to gastrointestinal, low back pain and anxiety disorders was 0.45.

$$\hat{\gamma}(CMPD\ Vs.\ GLAD) = \exp\{-0.7739\} = 0.46$$

$$\hat{\gamma}\,(CPD\ Vs.\ GLAD) = \exp\{-0.7992\} = 0.45$$

These results suggest that the median (or any other quartile) LOH until death was diminished by a factor of 0.46 for individuals with cardio-metabolic and pain disorders compared to those with gastrointestinal, low back pain and anxiety disorders. Similarly, among patients with cardio-pulmonary disorders compared to those with gastrointestinal, low back pain and anxiety disorders the median (or any other quartile) LOH until death was decreased by a factor of 0.45.

4.3.5 Weibull Survivor functions for the multimorbidity patterns

The probability of surviving longer than t days for hospitalized patients with cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders can be calculated by the following survival functions:

$$s(t, CMPD) = exp\{-(0.04282 * t)^{1.6940}\}$$

$$s(t, CMPD) = exp\{-(0.04392 * t)^{1.6940}\}$$

$$s(t, CMPD) = exp\{-(0.01975 * t)^{1.6940}\}$$

The graph of the estimated survivor functions of the Weibull model is shown in Figure 4.3. From this figure, cardio-metabolic and pain disorders, and cardio-pulmonary disorders are very close to each other and also, their survival probabilities decrease sharply with time than the gastrointestinal, low back pain and anxiety disorders. In other words, individuals with gastrointestinal, low back pain and anxiety disorders have longer survival days as compared to persons admitted with cardio-metabolic and pain disorders, and cardio-pulmonary disorders.

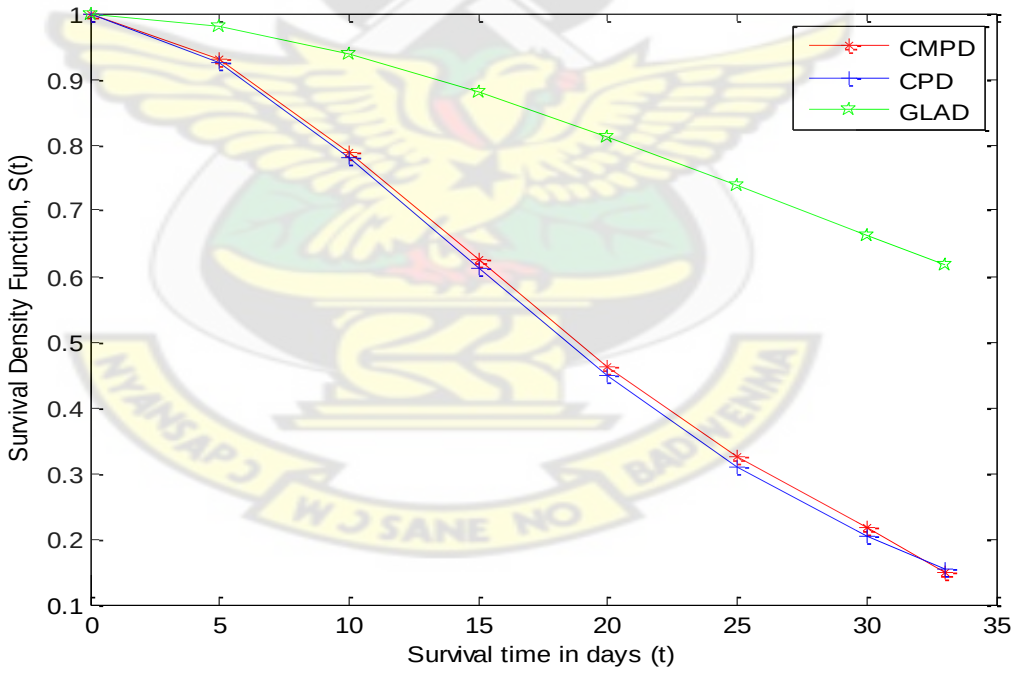

Figure 4.3: Weibull survivor curves for the multimorbidity patterns

Table 4.9: LOH until death estimated from the Weibull model

Quartile	S(t)	CMPD	CPD	GLAD
First quartile	0.25	32.4	31.6	70.2
Second quartile (Median)	0.50	16.2	15.8	35.1
Third quartile	0.75	6.7	6.6	14.6

Table 4.9 depicts the estimated length of hospitalization until death for the three multimorbidity patterns. Among patients with cardio-metabolic and pain disorders, and patients with cardio-pulmonary disorders their median length of hospitalization until death were approximately 16 days. Also the median length of hospitalization until death for patients with gastrointestinal, low back pain and anxiety disorders was 35 days. Results for the first and third quartile for CMPD and CPD were also similar. However, it is evident that the impact of CMPD and CPD on the LOH until death were almost the same.

4.3.6 Weibull hazard functions for the Multimorbidity patterns

From Table 4.8, the Weibull shape parameter of 1.6940 suggests an increasing Weibull hazard function for all the three multimorbidity patterns. This is demonstrated graphically in Figure 4.5. The respective hazard functions for cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders are given below.

$$h(t, CMPD) = 1.6940 * 0.04282 * t^{1.6940-1}$$

$$h(t, CMPD) = 1.6940 * 0.04392 * t^{1.6940-1}$$

$$h(t, CMPD) = 1.6940 * 0.01975 * t^{1.6940-1}$$

The graph of the estimated hazard functions $h(t)$ of the Weibull model is shown in Figure 4.4. From this figure, the hazard probabilities increased with the length of hospitalization for all the patterns of multimorbidity. For the case of cardio-pulmonary disorders, and cardio-metabolic and pain disorders increasing hazard probabilities were substantial. The probability of mortality among patients with cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders were 0.38, 0.39 and 0.16 respectively on the 10th day of hospitalization. These figures increased to 0.58, 0.59 and 0.34 for these same conditions on the 20th day. These suggest that mortality was highly probable among patients with cardio-pulmonary disorders, followed by those with cardio-metabolic and pain disorders and lastly by patients with gastrointestinal, low back pain and anxiety disorders.

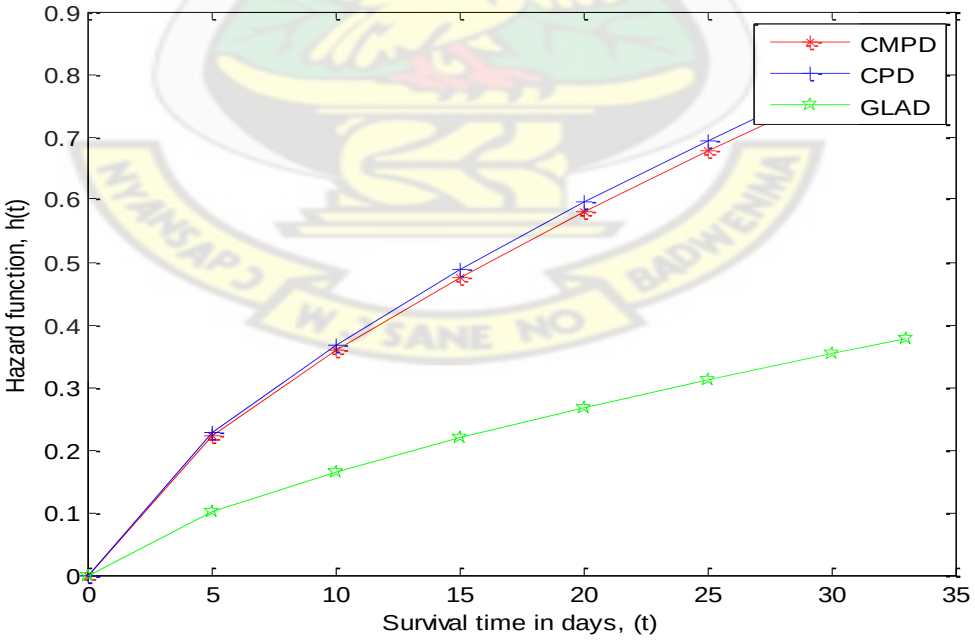

Figure 4.4: Weibull hazard curves for the multimorbidity patterns

Figure A2 in the appendix displays a graphical procedure for checking the adequacy of the Weibull model. The resulting lines are approximately straight, therefore suggesting that the Weibull assumption is reasonable. Moreover, the lines appear to have the same slope (i.e., are parallel), suggesting that both the PH and AFT assumptions hold for the Weibull model. These imply that our survival data follows the Weibull distribution and also the Weibull model reflects the effect of the multimorbidity patterns on the length of hospitalization. Therefore, these suggest that the Weibull model is adequate.

4.3.7 Choosing the appropriate parametric survival model

Table 4.10: Log likelihood and AIC statistic for the exponential and Weibull model

Model	Log likelihood	AIC statistic
Exponential model	-125.633	259.266
Weibull model	-115.827	**239.654**

Table 4.10 shows the log likelihood and the Akaike's information criterion (AIC) for the exponential and the Weibull model. A smaller AIC suggests a better fit. However, based on the AIC, the Weibull model is selected yielding the smallest AIC statistic at 239.654. Therefore, this suggests that the Weibull model provides a better fit than the exponential model. Consequently, the discussion and conclusion of the results of the present study was based on the findings of the Weibull model.

4.4 Discussion

In the present study, multimorbidity was defined as the co-existence of two or more chronic diseases in the same individual from a list of 13 chronic diseases. The overall prevalence of multimorbidity was estimated as high as 68.1% in the elderly general clinic patients of Kwadaso S.D.A hospital. This conforms to the reported range of prevalence of 3.5% to 98.5% in primary care (Fortin, et al. 2012). Regarding the pair-wise association of diseases, we found that some diseases jointly occurred together at higher chance. These were hypertension and diabetes (0.83), asthma and congestive heart failure (0.81), gastritis and low back pain (0.81), migraine and stroke (0.80), and asthma and stroke (0.75).

The co-existence of hypertension, diabetes mellitus, arthritis and migraine indicated the existence of cardio-metabolic and pain disorders. Secondly, the association of asthma, congestive heart failure and stroke revealed the existence of cardio-pulmonary disorders, and finally, the co-existence of gastritis, peptic ulcer disease, low back pain and anxiety indicated the existence of gastrointestinal, low back pain and anxiety disorders. The dominant among them was cardio-metabolic and pain disorders with total prevalence of 28.2%. The individual chronic diseases with prevalence exceeding this figure were hypertension (51.6%) and diabetes mellitus (36.3%). The prevalence of congestive heart failure (28.2%) was exactly the same as that of aforementioned multimorbidity pattern. Cardio-metabolic and pain disorder pattern was common in males (14.2%) than females (13.9%). For cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders their total prevalence were estimated as 13.9% and 10.1% respectively. Cardio-pulmonary disorder pattern was

frequent in females (8.6%) than males (5.3%). Similarly, gastrointestinal, low back pain and anxiety disorder pattern was also common in females (6.6%) than males (3.5%). Among the multimorbid patients, 15.9% could not be assigned to any of the three multimorbidity patterns. Moreover, the proportion of people with these multimorbidity patterns differed significantly for both sexes. However, their absolute prevalence was prevalent in females (29.2%) than males (23.0%).

The prevalence rates of cardio-metabolic and pain disorders were 2.7% in patients aged 50-54 years, 2.8% in those aged 55-59 years, 8.9% in 60-64 years and 13.6% in those aged 65 years or more. The prevalence of cardio-pulmonary disorders were estimated as 2.1%, 1.9%, 4.5% and 5.3% in patients aged 50-54, 55-59, 60-64 and 65 years or older respectively. In the same age groups the prevalence rates of gastrointestinal, low back pain anxiety disorders were estimated as 1.3%, 1.7%, 2.6% and 4.4%. Nevertheless, we found that the incidence of the identified multimorbidity patterns was the same across the age groups under discussion.

We demonstrated that our survival data follows both the exponential and the Weibull distributions. However, we found the Weibull model to be more useful and reliable than the exponential model based on the PH and AFT assumptions. Again the Weibull model with Akaike's information criterion of 239.654 provided a better fit when compared to the exponential model with Akaike's information criterion of 259.266. We found a significant association between the multimorbidity patterns and the log of the length of hospitalization until death. Cardio-pulmonary disorders, and cardio-metabolic and pain disorders showed an inverse association with parameter estimates -0.7992 and -0.7739 respectively.

The median (or any other quartile of) length of hospitalization was decreased by a factor of 0.46 for individuals with cardio-metabolic and pain disorders compared to those with gastrointestinal, low back pain and anxiety disorders. Also, among patients with cardio-pulmonary disorders compared to those with gastrointestinal, low back pain and anxiety disorders the median (or any other quartile of) length of hospitalization was decreased by a factor of 0.45.

The median length of hospitalization until death for patients with cardio-metabolic and pain disorders, and patients with cardio-pulmonary disorders were both estimated as 16 days and 35 days for patients with gastrointestinal, low back pain and anxiety disorders.

We found an increasing likelihood of mortality for all the multimorbidity patterns among the elderly patients who were hospitalized. This was substantial among patients with cardio-pulmonary disorders, followed by patients with cardio-metabolic and pain disorders.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

In this chapter, the conclusions were made based on the study findings and the recommendations were also made based on the conclusions drawn.

5.1 Conclusions

Multimorbidity is high among elderly general clinic patients of Kwadaso S.D.A hospital. The identified multimorbidity patterns: cardio-metabolic and pain disorders, cardio-pulmonary disorders, and gastrointestinal, low back pain and anxiety disorders affect a sizeable proportion of our sample. A large proportion of the sample variance was explained by these patterns (cumulative proportion: 0.8263). Moreover, the entire factor analysis model was acceptable (Kaiser's measure of sampling adequacy: 0.57).

Generally, the identified multimorbidity patterns appeared to be frequent in females than males. This may be due to the fact that females were considerably more than males in our sample. Moreover, the prevalence of all the multimorbidity patterns was higher among patients who were aged 65 years or older as compared to the age groups younger than this age. However, our chi-square test revealed that the existence of the multimorbidity patterns were the same across the ages groups.

We demonstrated that the length of hospitalization until death follows both the exponential and Weibull distribution. However we found that the accelerated failure time assumption only holds for the Weibull model, therefore making it appropriate for

the study. In addition, the Weibull model with AIC statistic of 239.654 provided a better fit when compared to the exponential model with AIC statistic of 259.266.

Among patients with cardio-metabolic and pain disorders, and those with cardio-pulmonary disorders their length of stay in the hospital until death were substantially influenced by their conditions.

The likelihood of mortality incrementally increased with the length of hospitalization among patients with cardio-metabolic and pain disorders, cardio-pulmonary disorders, and those with gastrointestinal, low back pain and anxiety disorders. Patients with cardio-pulmonary disorders and patients with cardio-metabolic and pain disorders were the most vulnerable. Therefore, mortality was highly probable in patients with cardio-pulmonary disorders, and those with cardio-metabolic and pain disorders.

5.2 Recommendations

More research on multimorbidity is needed particularly, in highly utilized hospitals in Ghana in order to monitor its prevalence and also ascertain how some chronic diseases cluster together so that appropriate control measures and strategies could be approved and implemented to control its prevalence.

Preventive and control measures of multimorbidity should focus on the entire elderly population and not specific ages of elderly people. Also attention must be focused equally among males and females.

Multidisciplinary and multi-professional team should be formed, based on the needs of each multimorbidity pattern or needs of each patient.

The identified multimorbidity patterns are clearly important health problem among elderly general clinic patients of Kwadaso S.D.A hospital that needs to be addressed. For that reason, medical professionals at Kwadaso S.D.A hospital should be more aware of the possible impact of these multimorbidity patterns, especially cardio-metabolic and pain disorders, and cardio-pulmonary disorders. Provision of care for these patients should be timely and done with utmost attention.

Further studies on more representative samples are required to confirm the existence of multimorbidity and specific multimorbidity patterns and their impact in the general population of Ghana. The interaction effects of different multimorbidity patterns should be taken into consideration.

REFERENCES

1. Agborsangaya, C.B, Lau, D, Lahtinen M, Cooke T, and Johnson J.A (2012). Multimorbidity prevalence and patterns across socioeconomic determinants: a cross-sectional survey. BMC Public Health 10:718.
2. Alvin, C. R. (2002). Methods of multivariate analysis. Second edition, John Wiley & Sons, Inc. publication, Canada.
3. Anderson, G, and Horvath, J. (2002). Chronic conditions: Making the case for ongoing care. Baltimore (MD): Partnership for solutions.
4. Barnett, K, Mercer, S.W, Norbury, M, Watt, G, Wyke, S, and Guthrie, B, (2012). Epidemiology of multimorbidity and implications for health care research and medical education: a cross sectional study. DOI: 10.1016/50140-6736(2)60240-2.
5. Bonnet, D.G, and Price, R.M. (2005). Inferential methods for the tetrachoric correlation coefficient. Journal of Education and Behavioral Statistics, 30:213-225.
6. Boyd, C.M, Darer, J, Boult, C, Fried L.P, Boult L, and Wu A.W, (2005). Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases. JAMA 294:716-24.
7. Boyd C.M and Fortin, M. (2010). Future of multimorbidity research: How should understanding of multimorbidity inform health system design? Public Health Reviews; 32: 451-74.
8. Britt, H.C, Harrison, C.M, Miller G.C and Knox, S.A (2008). Prevalence and patterns of multimorbidity in Australia. MJA. 189: 72-77.

9. Brown, M.B. (1977). Algorithm AS 116: The tetrachoric correlation and its asymptotic standard error. Journal of the Royal Statistical Society. 26:343-351.
10. Castellan, N.J. Jr. (1966). On the estimation of tetrachoric correlation coefficient. Psychometrika 31: 67-73.
11. Cornell, J.E, Pugh, J.A, Williams, Jr. J.W, Kazis, L, Lee, A. F.S, Parchman, M.L, Zeber, J, Pederson T, Montgomery, K.A and Nokl, P.H (2007). Multimorbidity clusters: Clustering binary data from a large administrative medical database. Applied Multivariate Research. 12:163-182.
12. Dobson, A. J.C. (2002). An introduction to generalized linear models. Second edition, Chapman & Hall/CRC, USA.
13. Feinstein, A.R. (1970). The pre-therapeutic classification of comorbidity in chronic disease. J Chron Dis; 23: 455-68
14. Fortin, M, Hudon C, Haggerty J, van den Akker, M and Almiral, J. (2010). Prevalence estimates of multimorbidity: a comparative study of two sources. BMC Health Services Research. 10.111
15. Fortin, M, Stewart, M, Poitras, M-E, Almiral, J and Maddocks, H (2012). A systematic review of prevalence studies on multimorbidity: Toward a more uniform methodology. Annals of family mediciene. 10(2): 142-151. Doi. 10.1370/afm.1337
16. Fortin, M, Bravo, G, Hudon, C, Vannasse, A, and Lapointe, L. (2005). Prevalence of multimorbidity among adults seen in family practice. Annals of family mediciene. 3: 223-228. Doi. 10.1370/afm. 272

17. Fortin, M, Lapointe, L, Hudon, C, Vannasse, A, Ntetu, A.L and Maltais, D. (2004). Multimorbidity and quality of life in primary care: a systematic review. Health and Quality of Life Outcomes. 2:51

18. Fortin, M., Lapointe, L., Hudon, C, and Vanasse, A. (2005). Multimorbidity is common to family practice. Can fam physician. 51:244-245

19. Gijsen, R., Hoeymanns, N., Schellevis, F.G., Ruwaard, D, and Satariano, W.A (2001). Causes and consequences of comorbidity: a review. J Clin Epidemiol 54: 661-674.

20. Guthrie, B., Payne K., Alderson, P., McMurdo, M.E.T, and Mercer, S.W. (2012). Adapting clinical guidelines to take account of multimorbidity. BMJ; 345e6341.

21. Guthrie, B., McCowan, C., Davey, P., Simpson, C.R., Dreischulte T., and Barnett, K. (2011). High risk prescribing in primary care patients particularly vulnerable to adverse drug events: cross-sectional population database analysis in Scottish general practice. BMJ; 342: d3514.

22. Guthrie, B., Payne K., Alderson, P., McMurdo, M.E.T., and Mercer, S.W. (2012). Adapting clinical guidelines to take account of multimorbidity. BMJ; 345e6341

23. Glynn, L.G., Valderas, J.M., Healy, P., Burke, E., Newell, J., Gillespie, P., and Murphy, A.W. (2011). The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. Family practice: 0: 1-8. doi: 10.1093/fampra/cmr013.

24. Hartman, J., Hehner, S., Hemmrich, K., Körs, B., and Möhlmann, T. (2011). Providing better care at lower cost for multimorbid patients. Health International; 11:39-47

25. Holden, L., Scuffham, P.A., Hilton, M.F., Muspratt, A., Shu-Kay, Ng., and Whiteford, J.A. (2011). Patterns of multimorbidity in working Australians. Population Health Metrics 9: 1-5.

26. Hudon, C., Soubhi, H., and Fortin, M. (2008). Relationship between multimorbidity and physical activity: Secondary analysis from the Quebec health survey. BMC Public Health; 8:304

27. Hunger, M., Thorand, B., Schunk, M., Döring, A., Menn, P., Peters, A., and Holle, R. (2011). Multimorbidity and health-related quality of life in the older population: results from the German KORA-Age study. Health and Quality of Life Outcomes. 9:53

28. John, R., Kerby, D.S., and Hennessy, C.H. (2003). Patterns and impact of comorbidity and multimorbidity among community-residents of American Indian elders. The Gerontologist. 43: 649-660.

29. Johnson, R.A and Dean W.W (2007). Applied Multivariate Statistical Analysis. Sixth Edition, Pearson Education International, U.S.A.

30. Juras, J., and Pasarić, Z. (2006). Application of tetrachoric and polychoric correlation coefficients to forest verification. Geofizika 23: 59-82

31. Kadam, U.T., and Croft, P.R. (2007). North Staffordshire GP Consortium Group. Clinical multimorbidity and physical function in older adults: a record and health linkage study in general practice. Family practice; 24:412-419

32. Kaiser, H.F. (1970), "A second generation Little Jiffy, Psychometrika, 35, 401-415.
33. Kaiser, H.F., and Rice, J. (1974), "Little Jiffy, Mark IV," Educational and phsychological Measurement, 34,111-117.
34. Khanan, M.A., Streatfield, P.K., Kabir, Z.N., Qiu, C., Cornelious, C., and Wahlin, A. (2011). Prevalence and patterns of multimorbidity among elderly people in rural Bangladesh: A cross-sectional study. JHPN 29:406-41.
35. King M.R., and Nipa A. M (2010). Numerical and Statistical Methods for Bioengineering. Applications in MATLAB. Cambridge University Press, New York, USA.
36. Kirchberger, I, Meissenger, C., Heier, M., Zimmermann, A-K., Thorand, B., Autenrieth, C.S., Peters, A., Ladwig, K-H, and Döring, A. (2012). Patterns of Multimorbidity in the Aged Population. PloS ONE 7(1):e30556. Doi: 10.1337/journal.pone.0030556.
37. Kleinbaum, D.G and Klein, M (2005). Survival analysis. A self learning text. Second edition, Springer, NY, USA.
38. Kubinger, K.D. (2003). On artificial results due to using factor analysis for dichotomous variables. Psychology science 45:106-110.
39. Laux, G., Kuehlein, T., Rosemann, T., and Szecsenyi (2008). Co- and multimorbidity patterns in primary care based on episodes of care: results from German CONTENT project. BMC Health Service Research 8: 14.
40. Long, M.A., Berry K.J., and Mielke P.W. Jr. (2009). Tetrachoric correlation: A Permutation Alternative. SAGE publications. 66: 429-437.

41. Mangin, D.I., and Jamoulle, M. (2012). Beyond diagnosis: rising to multimorbidity challenge. BMJ; 344:e3526.
42. Marengoni, A. (2008). Prevalence and impact of chronic diseases and multimorbidity in the aging population: A clinical and epidemiological approach. Aging Research Center (ARC), Karolinska Institutet, Stockholm, Sweden.
43. Marengoni, A., Winblad, B., Karp, A., and Fratiglioni, L. (2008). Prevalence of chronic diseases and multimorbidity among elderly population in Sweden. American Journal of Public Health; 98: 1198-1200.
44. Marengoni, A., Angleman, S., and Fratiglioni, L. (2011). Prevalence of disability according to multimorbidity and disease clustering: a population based study. Journal of Comorbidity 1:11-18.
45. Nagel, G., Peter, R., Braig, S., Hermann, S., Rohrmann, S., and Linseisen, J. (2008). The impact of education on risk factors and the occurrence of multimorbidity in the EPIC-Heildelberg cohort. BMC Public Health; 8:384.
46. Nobili, A., Garattini, S., and Mannucci, P.M. (2011). Multiple diseases and polypharmacy in the elderly: challenges for the internist of the third millennium. Journal of Comorbidity; 1:28-44.
47. Pearson, K. (1900). Mathematical contribution to the theory of evolution. VII. On correlation of characters not quantitatively measurable. Philosophical Transactions of the Royal Society of London, 195A, 1-47.
48. Pearson, K. (1913). On the probable error of a coefficient of correlation as found from a fourfold table. Biometrica, 19, 22-27.

49. Prados-Torres, A., Poblador-Plou B., Calderón-Larrañaga, A, Gimeno-Feliu, L.A., González-Rubio, F, Poncel-Falco, A., Sicras, Mainar, A., and Alcalá-Nalvaniz, J.T. (2012). Multimorbidity patterns in primary care: Interraction among chronic diseases using Factor Analysis. PLos ONE 7(2): e32190. doi: 10:1371/journal.pone 0032190.
50. Schäfer, I., von Leithner, E-C., Schön, G., Koller, D., Hansen, H., Kolonko, T., Kaduszkiewwicz, H, Wegscheide K, Glaeske, G., and Busche van den H. (2010). Multimorbidity Patterns in the Elderly. A new approach of Disease clustering Identifies Complex Interrelations between Chronic Conditions. PLoS ONE 5(12): e15941 doi: 10:1371/journal.pone: 0015941.
51. Schäfer, I., Hasen, H., Schön, G., Höfels, S., Altiner, A., Dahlhaus, A., Gensichen, J., Riedel-Heller, S., Weyerer, S., Blank, W.A, König, H-H., von dem Knesebeck, O., Wegscheider, K., Schere, M., van den Bussche, H., and Wiesse, B. (2012). The influence of age, gender and socio-economic status on multimorbidity patterns in primary care. First results from the multicare cohort study. BMC Health Service Research. 12.89.
52. Smith, S.M., Soubhi, H., Fortin, M., Hudon, C., and O'Dowd, T. (2012). Managing patients with multimorbidity: Systematic review of interventions in primary care and community settings. BMJ 2012; 345:e5205 doi: 10.1136/bmj.e5205.
53. Taylor, A.W., Price, K., Gill, T.K., Adams, R., Pilkington, R., Carrangis, N., Shi, Z., and Wilson, D. (2010). Multimorbidity-not just an older person's issue. Results from an Australian biomedical study. BMC Public Health 10:718.

54. Tucker-Seeley, R.D., Li, Y., Sorensen, G., and Subramanian, S.V. (2011). Lifecourse socioeconomic circumstances and multimorbidity among older adults. BMC Public Health; 11:313.

55. United Nations, Department of Economics and Social Affairs, Population Division (2012). Population Ageing and Development 2012. http://www.unpopulation.org.

56. United Nations, Department of Economics and Social Affairs, Population Division (2011). World Population Prospects: The 2010 revision, vol. 1: comprehensive tables. ST/ESA/SER.A/313.

57. van den Akker, M., Buntinx, F., and Knottnerous, J.A. (1996). Comorbidity or multimorbidity: What's in a name? A review of literature. Eur J Gen. Pract 2: 65-70.

58. van den Bussche, H., Koller, D., Kolonko, T., Hansen, H., Wegscheider, K., Glaeske, G., von Leitner, E.C., Schäfer, I., and Schön, G., (2011). Which chronic disease and disease combinations are specific to multimorbidity in the elderly? Results of a claims data based cross-sectional study in Germany. BMC Public Health 11:101.

59. van den Bussche, Schön, G., Kolonko, T., Wegscheider, K., Glaeske, G., and Koller, D. (2011). Patterns of ambulatory medical care utilization in elderly patients with special reference to chronic diseases and multimorbidity- Results from a claims data based observational study in Germany. BMC Geriatrics 11:54.

60. van Oostrom, S.H., Picavet, H.S.J., van Gelder, B.M., Lemmens, L.C., Hoeymans, N., van Dijk, C.E., Verheij, R.A., Schevellis, F.G., and Baan, C.A. (2012). Multimorbidity and comorbidity in the Dutch population-data from general practices. BMC 12:715.

61. Valderas, J.M., Mercer, S.W. and Fortin, M. (2011). Research on patients with multiple health conditions: different constructs, different views, one voice. 1: 1-3.

62. Vyas, A., yun Pan, X., and Sambamoorthi, U. (2012). Chronic condition clusters and polypharmacy among adults. International Journal of Family Medicine; 2012:8.

APPENDICES

Table A1: Demographic characteristics of the elderly patients (n=984)

			Age				Total
			50-54	55-59	60-64	65+	
Sex	Male	Count	83	126	82	125	416
		% within Age	32.5%	59.7%	34.5%	44.6%	42.3%
	Female	Count	172	85	156	155	568
		% within Age	67.5%	40.3%	65.5%	55.4%	57.7%
Total		Count	255	211	238	280	984
		% within Age	100.0%	100.0%	100.0%	100.0%	100.0%

Table A2: Number of chronic diseases in patients

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	One chronic condition	314	31.9	31.9	31.9
	Two chronic conditions	295	30.0	30.0	61.9
	Three chronic conditions	249	25.3	25.3	87.2
	Four chronic conditions	88	8.9	8.9	96.1
	Five chronic conditions or more	38	3.9	3.9	100.0
	Total	984	100.0	100.0	

Table A3: Tetrachoric correlations matrix:

	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13
Var1	1	0.302	0.306	-0.287	-0.120	0.034	0.153	-0.191	0.399	-0.320	0.297	-0.311	0.266
Var2	0.302	1	0.213	-0.141	-0.188	0.814	0.133	0.151	0.434	-0.209	0.751	-0.272	0.346
Var3	0.306	0.213	1	-0.210	-0.270	0.041	-0.135	0.221	0.332	-0.264	0.378	-0.307	0.252
Var4	-0.287	-0.141	-0.210	1	0.229	0.226	-0.060	0.025	-0.550	0.830	-0.383	0.330	-0.429
Var5	-0.120	-0.188	-0.270	0.229	1	-0.084	0.138	-0.138	-0.460	0.202	-0.304	0.812	-0.466
Var6	0.034	0.814	0.041	0.226	-0.084	1	0.337	0.022	0.026	0.266	0.430	-0.113	0.176
Var7	0.153	0.133	-0.135	-0.060	0.138	0.337	1	-0.319	-0.212	0.115	-0.079	0.153	-0.104
Var8	-0.191	0.151	0.221	0.025	-0.138	0.022	-0.319	1	0.378	-0.086	0.319	-0.185	0.199
Var9	0.399	0.434	0.332	-0.550	-0.460	0.026	-0.212	0.378	1	-0.602	0.800	-0.467	0.519
Var10	-0.320	-0.209	-0.264	0.830	0.202	0.266	0.115	-0.086	-0.602	1	-0.480	0.330	-0.292
Var11	0.297	0.751	0.378	-0.383	-0.304	0.430	-0.079	0.319	0.800	-0.480	1	-0.355	0.427
Var12	-0.311	-0.272	-0.307	0.330	0.812	-0.113	0.153	-0.185	-0.467	0.330	-0.355	1	-0.495
Var13	0.266	0.346	0.252	-0.429	-0.466	0.176	-0.104	0.199	0.519	-0.292	0.427	-0.495	1

Factor Analysis with Oblique (Oblimin) Rotation

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.56882417

Table A4: Eigenvalues of the Reduced Correlation Matrix: Total = 9.22971333 Average = 0.70997795

	Eigenvalue	Difference	Proportion	Cumulative
1	4.38525301	2.40494339	0.4751	0.4751
2	1.98030961	0.71950031	0.2146	0.6897
3	1.26080930	0.27769174	0.1366	0.8263
4	0.98311756	0.49214576	0.1065	0.9328
5	0.49097181	0.29160384	0.0532	0.9860
6	0.19936796	0.03956447	0.0216	1.0076
7	0.15980349	0.02605816	0.0173	1.0249
8	0.13374533	0.08587016	0.0145	1.0394
9	0.04787517	0.09016959	0.0052	1.0446
10	-.04229442	0.02598174	-0.0046	1.0400
11	-.06827617	0.03905949	-0.0074	1.0326
12	-.10733566	0.08629802	-0.0116	1.0210
13	-.19363367		-0.0210	1.0000

Factor Analysis with Oblique (Oblimin) Rotation

The FACTOR Procedure
Initial Factor Method: Principal Factors

3 factors will be retained by the MINEIGEN criterion.
Factor Analysis with Oblique (Oblimin) Rotation

Table A5: Initial Factor Method: Principal Factors

Factor Pattern

		Factor1	Factor2	Factor3
Var1	Arthritis	0.45067	-0.01033	0.36493
Var2	Asthma	0.63065	0.69288	0.14985
Var3	Liver cirrhosis/Hepatoma	0.44056	-0.04614	-0.07730
Var4	Diabetes	-0.61557	0.43945	-0.42666
Var5	Gastritis	-0.58962	0.09056	0.48372
Var6	Congestive heart failure	0.23545	0.93491	0.00715
Var7	Peptic ulcer disease	-0.11741	0.32356	0.44720
Var8	Anxiety	0.28545	-0.00861	-0.42194
Var9	Migraine	0.86662	-0.14147	-0.01443
Var10	Hypertension	-0.65391	0.46041	-0.38062
Var11	Stroke	0.82792	0.27599	0.07393

Var12	Low back pain	-0.67879	0.09792	0.37458
Var13	Kidney stones	0.62308	-0.00982	-0.11942

Variance Explained by Each Factor

Factor1	Factor2	Factor3
4.3852530	1.9803096	1.2608093

Table A6: Final Communality Estimates: Total = 7.626372

Var1	Var2	Var3	Var4	Var5	Var6	Var7
0.33638692	0.90024608	0.20219765	0.75408084	0.58983324	0.92955452	0.31846428

Var8	Var9	Var10	Var11	Var12	Var13
0.25959064	0.77124727	0.78444731	0.76708467	0.61064993	0.40258858

Table A7: Root Mean Square Off-Diagonal Partials: Overall = 0.22835453

Var1	Var2	Var3	Var4	Var5	Var6	Var7
0.22062843	0.19910233	0.09107206	0.19981821	0.28184372	0.22981955	0.21012501

Var8	Var9	Var10	Var11	Var12	Var13
0.22301364	0.26892082	0.20135945	0.29905878	0.30883454	0.13242268

The FACTOR Procedure
Rotation Method: Oblimin (tau = 0)

Oblique Transformation Matrix

	1	2	3
1	-0.58813	0.33470	-0.50007
2	0.50866	0.96359	0.13521
3	-0.72391	0.11490	0.90704

Inter-Factor Correlations

	Factor1	Factor2	Factor3
Factor1	1.00000	-0.21740	0.28259
Factor2	-0.21740	1.00000	-0.12006
Factor3	0.28259	-0.12006	1.00000

Table A8: Rotated Factor Pattern (Standardized Regression Coefficients)

		Factor1	Factor2	Factor3
Var1	Arthritis	-0.53448	0.18282	0.10424
Var2	Asthma	-0.12694	0.89595	-0.08577
Var3	Liver cirrhosis/Hepatoma	-0.22662	0.09412	-0.29666
Var4	Diabetes	0.89443	0.16839	-0.01976
Var5	Gastritis	0.04267	-0.05451	0.74584
Var6	Congestive heart failure	0.33190	0.98051	0.01515
Var7	Peptic ulcer disease	-0.09010	0.32387	0.50809
Var8	Anxiety	0.13318	0.03876	-0.52663
Var9	Migraine	-0.57120	0.15208	-0.46558
Var10	Hypertension	0.89431	0.18105	0.04402
Var11	Stroke	-0.40005	0.55155	-0.30965
Var12	Low back pain	0.17786	-0.08980	0.69244

Var13	Kidney stones	-0.28500	0.18536	-0.42123

Reference Axis Correlations

	Factor1	Factor2	Factor3
Factor1	1.00000	0.19267	-0.26469
Factor2	0.19267	1.00000	0.06261
Factor3	-0.26469	0.06261	1.00000

Factor Analysis with Oblique (Oblimin) Rotation

The FACTOR Procedure
Rotation Method: Oblimin (tau = 0)

Reference Structure (Semipartial Correlations)

		Factor1	Factor2	Factor3
Var1	Arthritis	-0.50309	0.17810	0.09979
Var2	Asthma	-0.11948	0.87280	-0.08211
Var3	Liver cirrhosis/Hepatoma	-0.21331	0.09169	-0.28401
Var4	Diabetes	0.84190	0.16404	-0.01891
Var5	Gastritis	0.04016	-0.05310	0.71404
Var6	Congestive heart failure	0.31241	0.95518	0.01451
Var7	Peptic ulcer disease	-0.08481	0.31550	0.48642
Var8	Anxiety	0.12536	0.03776	-0.50417
Var9	Migraine	-0.53765	0.14816	-0.44573
Var10	Hypertension	0.84179	0.17638	0.04214
Var11	Stroke	-0.37656	0.53730	-0.29644
Var12	Low back pain	0.16742	-0.08748	0.66291
Var13	Kidney stones	-0.26827	0.18057	-0.40327

Variance Explained by Each Factor Eliminating Other Factors

Factor1	Factor2	Factor3
2.3832673	2.2269725	1.9889817

Factor Structure (Correlations)

		Factor1	Factor2	Factor3
Var1	Arthritis	-0.54477	0.28651	-0.06875
Var2	Asthma	-0.34596	0.93384	-0.22920
Var3	Liver cirrhosis/Hepatoma	-0.33092	0.17900	-0.37200
Var4	Diabetes	0.85224	-0.02369	0.21279
Var5	Gastritis	0.26529	-0.15333	0.76445
Var6	Congestive heart failure	0.12302	0.90653	-0.00877
Var7	Peptic ulcer disease	-0.01693	0.28245	0.44375
Var8	Anxiety	-0.02406	0.07303	-0.49364
Var9	Migraine	-0.73583	0.33216	-0.64526
Var10	Hypertension	0.86739	-0.01866	0.27501
Var11	Stroke	-0.60747	0.67569	-0.48891
Var12	Low back pain	0.39306	-0.21160	0.75348
Var13	Kidney stones	-0.44434	0.29789	-0.52402

Factor Analysis with Oblique (Oblimin) Rotation
The FACTOR Procedure
Rotation Method: Oblimin (tau = 0)

Variance Explained by Each Factor Ignoring Other Factors

Factor1	Factor2	Factor3
3.3534304	2.6179220	2.8393289

Final Communality Estimates: Total = 7.626372

Var1	Var2	Var3	Var4	Var5	Var6	Var7
0.33638692	0.90024608	0.20219765	0.75408084	0.58983324	0.92955452	0.31846428

Var8	Var9	Var10	Var11	Var12	Var13
0.25959064	0.77124727	0.78444731	0.76708467	0.61064993	0.40258858

Model Information

Data Set	WORK.MULTIMORBIDITY
Dependent Variable	Log(days)
Censoring Variable	censor
Censoring Value(s)	1
Number of Observations	154
Noncensored Values	61
Right Censored Values	93
Left Censored Values	0
Interval Censored Values	0
Name of Distribution	Exponential
Log Likelihood	-125.6325764

Number of Observations Read	154
Number of Observations Used	154

Class Level Information

Name	Levels	Values
pattern	3	CMPD CPD GLAD

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
pattern	2	12.9968	0.0015

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.5827	0.3333	3.9294	5.2360	189.01	<.0001
pattern	CMPD	1	-1.2811	0.3773	-2.0206	-0.5416	11.53	0.0007
pattern	CPD	1	-1.3304	0.4014	-2.1171	-0.5437	10.99	0.0009
pattern	GLAD	0	0.0000	.	.	.	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		
Weibull Shape		0	1.0000	0.0000	1.0000	1.0000		

Model Information

Data Set	WORK.MULTIMORBIDITY
Dependent Variable	Log(days)
Censoring Variable	censor
Censoring Value(s)	1
Number of Observations	154
Noncensored Values	61
Right Censored Values	93
Left Censored Values	0
Interval Censored Values	0
Name of Distribution	Weibull
Log Likelihood	-115.8268365

Number of Observations Read	154
Number of Observations Used	154

Class Level Information

Name	Levels	Values
pattern	3	CMPD CPD GLAD

Algorithm converged.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
pattern	2	11.8350	0.0027

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	3.9246	0.2189	3.4956	4.3537	321.48	<.0001
pattern	CMPD	1	-0.7739	0.2363	-1.2371	-0.3107	10.72	0.0011
pattern	CPD	1	-0.7992	0.2513	-1.2916	-0.3067	10.12	0.0015
pattern	GLAD	0	0.0000	.	.	.	.	.
Scale		1	0.5903	0.0634	0.4782	0.7287		
Weibull Shape		1	1.6940	0.1820	1.3724	2.0911		

Reparameterized values of the parameter ϕ in the exponential model

$$\hat{\phi}(CMPD) = \exp\{-4.5827 + 1.2811\} = 0.0368$$

$$\hat{\phi}(CPD) = \exp\{-4.5827 + 1.3304\} = 0.0387$$

$$\hat{\phi}(GLAD) = \exp\{-4.5827\} = 0.0102$$

Reparameterized values of the parameter ϕ in the Weibull model

$$\hat{\phi}(CMPD) = \exp\{-3.9246 + 0.7739\} = 0.04282$$

$$\hat{\phi}\,(CPD) = \exp\{-3.9246 + 0.7992\} = 0.04392$$

$$\hat{\phi}\,(GLAD) = \exp\{-3.9246\} = 0.01975$$

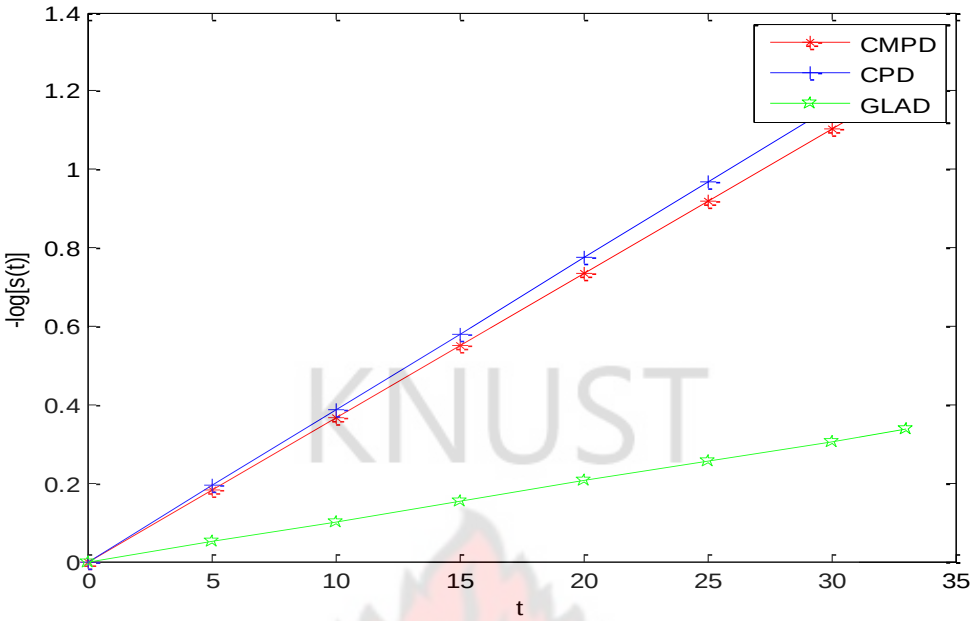

Figure A1: Exponential, PH and AFT assumptions of the exponential model

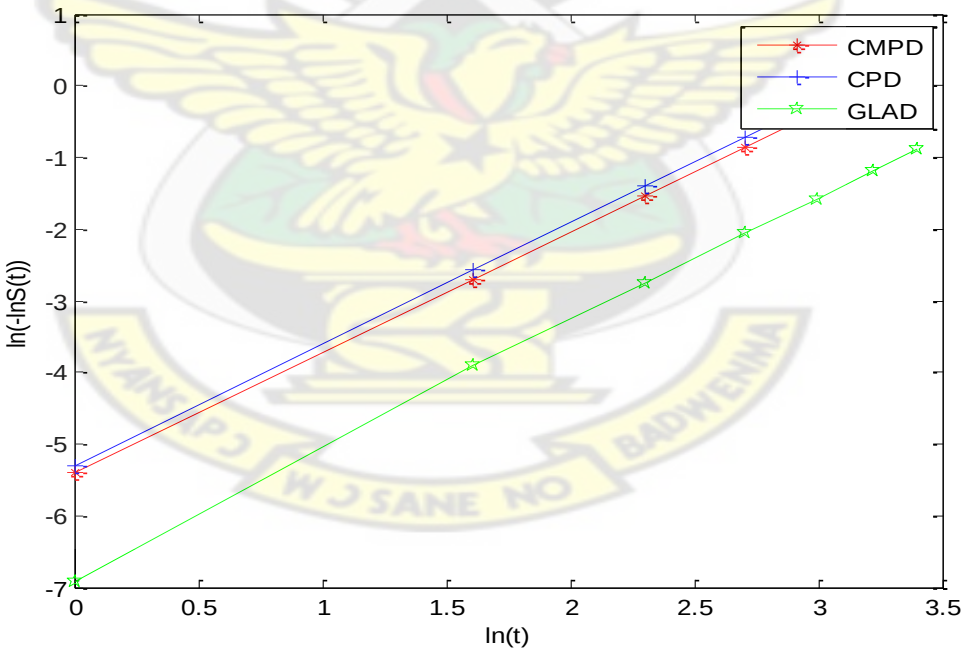

Figure A2: Weibull, PH and AFT assumptions of the Weibull model