INFLUENCE OF SOCIAL ACTIVITY ON ELDERLY (AGED 65+) SURVIVAL IN IDIKAN, IBADAN, OYO STATE, NIGERIA

BY

OSUNTOKUN TOLULOPE O.
MATRIC NO: 100589

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN MEDICAL STATISTICS

DEPARTMENT OF MEDICAL STATISTICS AND ENVIRONMENTAL HEALTH, FACULTY OF PUBLIC HEALTH, COLLEGE OF MEDICINE, UNIVERSITY OF IBADAN,

JUNE 2012.

DEDICATION

This project is dedicated to God

And

To the loving memory of my beloved Father,
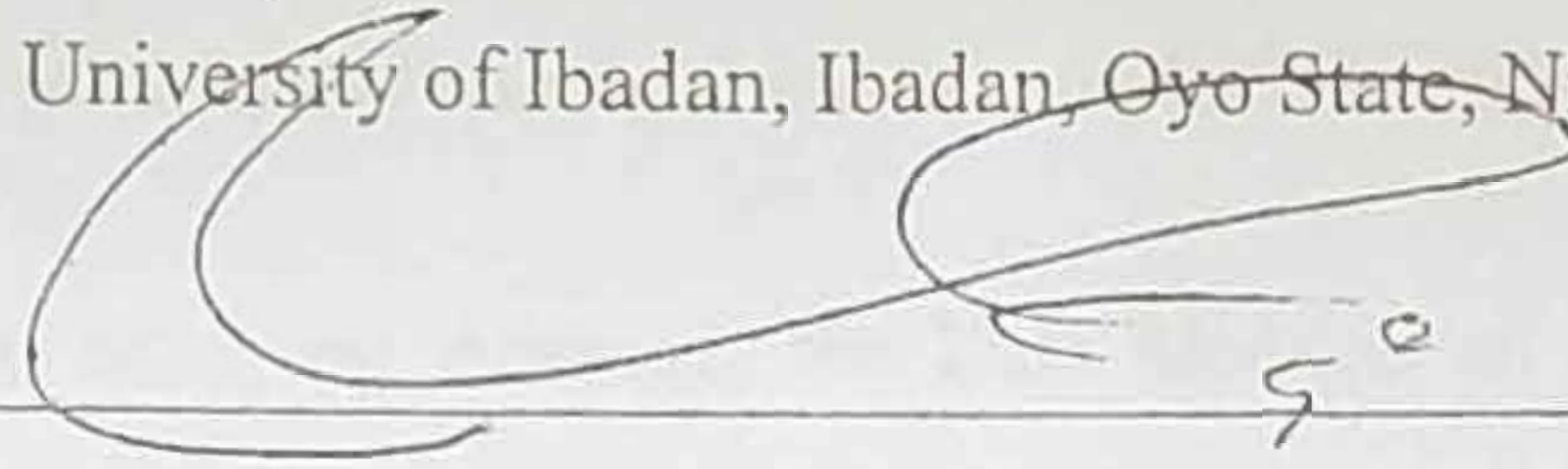
Chief Tomori Moses Olawobi Osuntokun.

CERTIFICATION

I hereby certify that this project was independently carried out by Osuntokun Tolulope O. under

my supervision in the Department of Epidemiology and Medical Statistics College of Medicine,

University of Ibadan, Ibadan, Oyo State, Nigeria.

SUPERVISOR

PROF. E.A. BAMGBOYE

B.Sc. (Lagos); M.Sc. (Med. Stat) (Lond.);

Ph.D. (Lond.); M.L.S. (UK); F.S.S. (Lond.)

iii

ACKNOWLEDGEMENT

I thank God almighty for his favour in the completion of this project. Special thanks to my supervisor, Prof. E.A Bamgboye, who during the process was appointed the Deputy Vice Chancellor of University of Ibadan (2008-2012). He found time in spite of his tight schedule to supervise the work and encouraged me to complete the dissertation.

I am grateful to Mr.Remi Adetunji, the Data Manager of Ibadan Indianapolis Dementia Research Project for his assistance in retrieving data used, from the project database. I thank the project Director of Ibadan-Indianapolis Dementia Research Project for allowing me to use their data and also for their moral and material support needed for the accomplishment of the project. I am also grateful to Mr. Odunayo Akinyemi, one of my lecturers in Medical Statistics and Environmental Health, MSEH department for his technical support during the data analysis. I give gratitude to the entire lecturers of MSEH department for putting me through various courses and impacting knowledge unto me.

I cannot but thank my immediate and extended family for their financial and moral support. I thank all those who had contributed in one way or the other to the success of my M.Sc. program especially my husband, Mr. Akin Akinrogunde. God bless you all.

ACKNOWLEDGEMENT

I thank God almighty for his favour in the completion of this project. Special thanks to my supervisor, Prof. E.A Bamgboye, who during the process was appointed the Deputy Vice Chancellor of University of Ibadan (2008-2012). He found time in spite of his tight schedule to supervise the work and encouraged me to complete the dissertation.

I am grateful to Mr.Remi Adetunji, the Data Manager of Ibadan Indianapolis Dementia Research Project for his assistance in retrieving data used, from the project database. I thank the project Director of Ibadan-Indianapolis Dementia Research Project for allowing me to use their data and also for their moral and material support needed for the accomplishment of the project. I am also grateful to Mr. Odunayo Akinyemi, one of my lecturers in Medical Statistics and Environmental Health, MSEH department for his technical support during the data analysis. I give gratitude to the entire lecturers of MSEH department for putting me through various courses and impacting knowledge unto me.

I cannot but thank my immediate and extended family for their financial and moral support. I thank all those who had contributed in one way or the other to the success of my M.Sc. program especially my husband, Mr. Akin Akinrogunde. God bless you all.

ABSTRACT

Back ground: There was a dearth of information on the relationship between social activities and survival of Nigerians at older age. This study provides some knowledge on the influence of social activity on the longevity of some Nigerians in Oyo State.

Objective: The main purpose of the study was to identify the relationship between social activities and survival of the aged (65+) in Idikan area of Ibadan, Oyo state, Nigeria, after adjusting for potential confounders.

Methods: This was a retrospective study, and a secondary analysis of data generated from a larger longitudinal prospective study. The study population consisted of all the elderly persons, aged 65 years and above living in Idikan area of Ibadan, Oyo State, Nigeria who were recruited into the Ibadan-Indianapolis Dementia Research Project at the initial stage in 1992/1993. The sampling method adopted was a total sampling of the entire study population Detailed information on the subjects' socio demographic background, life style and health history were collected at baseline in 1992/1993 with mortality follow-up of 17 years. Simple descriptive statistics were used to summarize the data and appropriate test statistics used to investigate relationship between variables. Kaplan Meier Curve and Cox regression model were used to describe and model survival from the start of the study to the end, examining the influence of the variables collected on survival and adjusting for potential confounders with p-value ≤ 0.05 and a 95% confidence interval. SPSS 11.0 for Windows was used to run the entire analyses.

Results: The studied population size was N= 2, 485. About 9% of the population was of age 85 years and above with majority (72.1%) between the age of 65–74 years and a male to female ratio of 1:2. A higher proportion of those who had impaired cognition, not socially active and males had died including those ever smoke and drink. The overall median survival

v

time for the 17 years mortality follow up period was estimated to be 8 years with the probability of surviving past it, approximately half. Kaplan Meir Curve showed that those who were not socially active presented with a poor survival than their counterpart and this was statistically significant at 5% levels. Cox regression model result showed that participation in social activities, gender, age, literacy status, presence of chronic disease (s) and cognitive status had influence on survival except for living arrangement, alcohol consumption and smoking that were not significant at $p \leq 0.05$. The study showed that those who were socially active were almost one and a half times more likely to live longer than those who were not socially active with p-value of 0.000 and a 95% C.I of 1.358 to 1.743.

Conclusion: Participation in social activity was found to be preventive measure against early death. It contributed considerably to achieving longevity and these highlights the need to consider formulating programs surrounded with social activities for elderly people in the society, aimed at promoting longer life.

TABLE OF CONTENTS

ix

LIST OF TABLES

LIST OF FIGURE

x

CHAPTER ONE

1.0. INTRODUCTION

The Nigeria life expectancy; which was 53 years in 1991(WHO, 1991), was observed to have declined and over the years to 48.2 years in 2003 (NDHS, 2003) and then increased to 48.4 years in 2010 (UN Report, 2010). This is in contrary to the expectation that it would rise from 53 years in 1991 to 57 years in 2003 and 62 years in 2013 (WHO, 1991) at the pace reported in 1991. However, it is estimated that with the expanding epidemics, life expectancy in the country will continue to decline.

In addition to communicable and infectious diseases, the morbidity and mortality rates associated with non-communicable diseases have also been on the increase in the past couple of decades and are set to wipe out all the gains and improvements through the primary health care delivery strategy adopted in the last two to three decades. Prominent among these non-communicable diseases are cardiovascular diseases, diabetes, stroke and various cancers. They are largely responsible for the progressive decline of life expectancy in the country. Reversal of these factors will contribute significantly to longer life expectancy in Nigeria.

Understanding the determinants of late-life survival in Nigeria becomes increasingly important in view of the current life expectancy declines. Therefore, the public health importance of factors that can contribute to the understanding and documentation of longer life expectancy in Nigeria cannot be over-emphasized. Since this will enable us to identify specific intervention or reliable strategies.

There are reports suggesting that participation in social activities may create a sense of belonging and emotional closeness, and are generally associated with a longer life (Petral L. Klumb, Heiner Maier 2007). In addition to physical influence of social activities, psychosocial effects are also affected. Involvement in social activities is said to have two

1

consequences, there is a direct effect on human being (Clark and Watson 1988) and a stress-buffering effect (House et al. 1988). The protective effect of social activities on the mortality of the elderly population has been shown in research using variables related to specific aspects of social activities (Hanson BS et al, 1989; Sabin EP et al, 1993, Yasuda N et al, 1997; Brown SL, 2003; Giles LC, 2005].

Recently more optimistic views of old age have been proposed by those focusing on "successful aging" (Ross et al., 1984) who suggest that environmental and physiologic modifiers of biological aging can be identified which may be amenable to preventive strategies even among the oldest-old. Most results on this topic are derived from studies conducted in Northern Europe, North America, and East Asia (Angel Rodriguez-Laso, 2007). There is a paucity of studies on factors that are related to late survival in Africa developing countries, Nigeria in particular. Research on social activities and mortality in Nigeria is also largely unavailable. Fortunately, the Ibadan-Indianapolis Dementia Research project has a reliable data bank from their prospective study. Records are available on the social activities of the elderly and this is yet to be analyzed and reported. An analysis of the data will add to the knowledge of relationship of social activities and survival of the elderly.

This dissertation provides information on the survival pattern in a seventeen (17) year (1992-2009) community based follow-up of elderly (aged 65+) in Idikan, Ibadan, Oyo state, Nigeria, with particular focus on the relationship between social activities and survival , adjusting for potential confounding effects of socio-demographic, life style and health conditions variables. The research is the analysis of a secondary data that were collected in a prospective study.

2

1.1. Main Objective

The main objective of this study is to identify (if any), the relationship between social activities and survival of the elderly (aged 65+) in idikan, Ibadan, Nigeria, over a seventeen (17) year period (1992-2009), after adjusting for potential confounders.

1.2. Specific Objectives:

- To describe the socio- demographic and health background of the elderly in Idikan, Ibadan, Oyo State, Nigeria.

- To describe the mortality pattern by socio - demographic characteristics and health status

- To estimate the yearly survival rate among the elderly in Idikan, Ibadan.

- To estimate and compare the survival curve of the two classes of elderly in Idikan based on their status of participation in social activities

- To identify the relationship between social activity and survival, adjusting for potential confounders such as socio-demographic (age, sex, education, live alone/do not live alone), life style- smoking, alcohol drinking, involvement in social activity-socially active, not socially active, health condition/history-presence of chronic disease (such as diabetes, cancer, stroke, cardiovascular disease) and cognitive status.

1.4. Research Hypothesis

The research hypothesis is that; there is relationship between involvement in social activities and time to death after adjusting for potential confounders such as socio-demographic, lifestyle and health condition.

3

CHAPTER 2

2.0. LITERATURE

Survival is the state of continuing to live or exist, often despite of difficulty or danger (Oxford Advanced Learners Dictionary, 7th edition). This pre-supposes one is alive at some point in time.

Baltes' two-component model broadly categorized all activities during waking day into two broad categories: regenerative and discretional (Baltes, Maas, Wilms, Borchelt & Little, 1999).

Regenerative activities are activities <u>that have to</u> be carried out by physiological necessity to maintain one's physical existence (e.g. personal hygiene, eating and resting) while discretional activities are activities that <u>can be</u> selected or that one can do on the basis of abilities and preferences.

Discretionary activities are further sub-divided into productive and consumptive activities (Reid, 1934). If an activity is performed predominantly due to its outcomes and can, therefore, be delegated to a third party without losing its benefit (e.g., doing laundry, house cleaning, running errands), then it is productive. In contrast, if an activity is performed primarily for its own sake and cannot, therefore, be delegated to a third party without losing benefit (e.g., meeting friends, reading a novel, watching TV), then it is consumptive.

Consumptive activities are also known as social activities: which are made up of active leisure, locomotion, health related activities, reading, watching TV/listening to radio, face-to-face talks, visiting, phone conversations, solving puzzles or reading a novel, playing games and other social interaction.

4

Participation in social activities provides social contacts and thereby fulfills a phylogenetically determined need for affiliation (Cantor & Sanderson, 1999; Reis, Sheldon, Gable, Roscoe, & Ryan, 2000). Engagement in such activities may create a sense of belonging, emotional closeness, learning and enjoyment. Involvement in this kind of activity has two consequences. There is a direct effect on well-being (Clark and Watson, 1988) and a stress-buffering effect (House et al., 1988). Activity theory (Lemon, Bengtson & Peterson, 1972; Longino & Kart, 1982) postulates that social activity is associated with life satisfaction because social activity provides opportunities for role support which in turn reaffirms the self-concept.

The influence of social participation on survival and health appears to be exerted by promoting psychobiological recovery processes which play a central role in the onset of age-related illnesses such as cardiovascular diseases, Type-II diabetes, and dementia (McEwen, 1998; Sapolsky, 1993). There are evidences that survival in the elderly is at least related to some level of social activities in which they are involved in. Previous studies primarily from developed countries have shown that participation in a social activity is related to long life (Evans et al., 1991; Heeren et al, 1992). A study of social integration of elderly in most intimate social circles, comprised of family and friends, as well as their social involvement in their communities protected against mortality (Rodriguez-Laso et al, 2007).

Social activities affect perceived quality of life (lawton et al, 1986-87) and they afford opportunities for aspiring, experience of competence, and autonomy. In contrast to some productive activities such as household chores, social activities are more challenging but the demands can be more easily adjusted to one's skill level or preferences because they are more freely chosen. Moreover, they yield more feedback and on average, more time is devoted to them per day. Engagement in social activities may thus affect mental and physical health and, ultimately, survival (Heiner Maier and Petra L. Klumb, 2007). Studies also posit that a

5

productive activity was not associated with survival in persons aged 70 and older where the majority of productive activities were household chores (Heiner Maier and Petra Klumb 2007). Thus, the multiple benefits, of participation in social activities are the promotion of physical and mental health, and ultimately, survival (Glass et al., 1999; Hendricks & Hendricks, 1990; House et al., 1982; Lennartson & Silverstein, 2001; Welin et al., 1992).

It is well known that there are some set of covariates that can determine the relationship between social activities and mortality; they are referred to as confounding factors or covariates. Survival rates have varied by severity of impairment, socio-demographic status (e.g. sex, age, and marital status), Life style (e.g. smoking, alcohol drinking, live alone) and health condition, such as presence of chronic disease (cancer, stroke, diabetes, cardiovascular problem), cognitive functioning and disabilities (Backlay et al, 1985; Heyman et al, 1987;Walsh et al, 1991; Burns et al; 1991, Kaszniak et al; 1978, Van Dijk et al, 1991). Previous studies, primarily from developed countries, have shown that dementia is associated with increased mortality with relative risk s ranging from 1.9 to 3.6 (Evans et al, 1991; Katzman et al, 1994; Jagger et al, 1995; Batdereschi et al, 1999; Tandi et al, 1999)

Studies have shown that, participation in regenerative, productive and consumptive activities is associated with survival in persons aged 70 and older. After adjusting for confounding influences, only Consumptive activities which dominates social activities were related to survival were related to survival (Relative Risk =0.76, 95% CI 0.58 to 1.00) (Heiner Maier and Petra Klumb, 2007)

Presence of chronic disease was associated to early death, decreasing cognitive functioning (with high probability of being demented) was associated with increased mortality, increasing age, being male sex, smoking were associated with increased risk of mortality or reduced survival while alcohol drinking was associated with increased risk of mortality(Anthony

6

Perkinser et al., 2002)

However, evidence based studies determining the association between social activities and survival in developing country has so far not available especially in Nigeria. Nevertheless, cultural variability may also explain differences in the associations between social activities and mortality as against the study findings shown in a study of three communities in the USA (Seeman, 1993).

CHAPTER 3

3.0. METHODOLOGY

3.1. STUDY AREA

The study area is Idikan, the North West local government area of Ibadan, the capital city of Oyo state with a population of over 3.5million located in the South Western part of Nigeria and adjacent contiguous eastern region of Benin. Idikan is an inner-core ancient part of Ibadan city, a densely populated area of Ibadan, with an estimated fairly stable population of 40,000 people, and a male to female ratio of 1:2 respectively. It comprises of Yoruba and mainly Muslim people of low socio-economic level. According to the baseline population study conducted at the start of the Ibadan/Indianapolis Dementia Research Project in 1992, 9% of the elderly population were aged 85years and above while 72% were aged between 65 to 74years and 19% were aged between 75 to 84 years. The resident of Idikan wards are typically small traders and craftsmen. Water supply is mainly from the well dug in the neighborhood and electricity supply is not consistent and often nonexistence.

3.2. STUDY TYPE

This is a retrospective study. The event – outcome variable (Death) and the explanatory variables (participation in social activities and other) have already occurred in the past. Data were merely collected now to establish the relationship between the outcome variable and the exposure variable. The study period is seventeen 17 years; study start in 1992 and end in 2009.

3.3. DATA SOURCE

It is a secondary data generated from a longitudinal prospective study initiated in 1992 by a group of research team of investigators from Indianapolis University and University of Ibadan which is refer to as Ibadan-Indianapolis Dementia Research Project. The main purpose of this larger longitudinal study was to describe the incidence and prevalence of

8

Dementia and Alzheimer disease (AD) among the Yoruba elderly community aged 65years and older, living in Idikan area of Ibadan city in Oyo state of Nigeria.

3.4. STUDY POPULATION

The population for this study comprises of all eligible elderly (aged 65years and above) in Idikan who were recruited into the study of Ibadan Dementia Research Project during the baseline population survey in 1992/1993.

3.5. SAMPLING METHOD

It was a total sampling of the entire eligible elderly (aged 65years and above) in Idikan who were recruited into the Ibadan-Indianapolis Dementia Research Project during the baseline population survey in 1992/1993. A total of 2,494 elderly were interviewed and recruited from the total of 3,489 households enumerated at the baseline survey making the sample size to be N=2,494.

3.6. DATA COLLECTION METHOD AND TOOL

A baseline cross sectional survey was conducted in 1992 at the start of the 'Ibadan-Dementia Research Project' to determine the prevalence of the Alzheimer disease among the total population of elderly (65years and above) in Idikan. Household enumeration and census in Idikan was completed for 3,489 households in preparation for the baseline study. Using a Quantitative method of data collection with a well-structured questionnaire called Community Screening Instrument for Dementia (CSI-D), interview were completed for 2,494 individuals by means of a door-to-door screening and a well-trained interviewers of which: 9% of the population were aged between 65-74 years, 19% were aged 75-84 years and 19% were aged between 75-84 years; sex ratio is 35% male and 65% female. The CSI-D consist of a written informed consent signed by all participants and basic information such as participant's level of participation in daily and social activities, presence of chronic disease,

9

lifestyle-smoking and alcohol drinking, lives alone/not live alone, socio-demographic status, disability status, cognitive status were collected. All the data collected at the baseline survey were entered into a computerized data base with FoxPro software and vital status and date of death were monitored regularly by the interviewers visit to the study site and the database was periodically updated.

Permission was obtained from the Ibadan Dementia Research Project Team to link with the computerized project database to collect the secondary data used for this study. The study period was between March, 1992 and March, 2009. In 2009, a seventeen follow-up year was undertaken for all persons in this study. Death status was ascertained for the entire subject in the cohort. The key variables implored in this study from the large study are socio-demographic variables such as age, sex, education, house composition-lives alone/not live alone; life style variables such as smoking status, alcohol drinking status, involvement in social activities status- socially active/non socially active; health history variables-presence/absence of any chronic diseases (Diabetes, Cancer, Stroke, Cardiovascular Disease) and cognitive status.

3.7. STATISTICAL METHOD AND ANALYSIS TOOL USED

It is a secondary analysis of the data generated from the Ibadan-Indianapolis Dementia Research Project. Survival analysis technique was the statistical method used. Out of the sample size of N=2, 494, expected to be involve in the analysis, only 9 individuals were excluded from the final analysis because of missing information resulting into the final population studied size of N=2,485. Frequency table was used to present subjects background. Logistic regression models were used to derive estimates of effect for analyses related to mortality or death. Kaplan –Meier Method was used to estimate the yearly survival probabilities, survival curve, described and compared the survival pattern, calculate the overall and the group median survival time for the two categories of the elderly in Idikan.

10

The log-rank test was used to whether the survivor curves are statistically different. Cox regression models were used to model time from screening to death or lost to follow up across the two groups of elderly as a function of activities involvement adjusting for covariates or confounding factors such as socio-demographic (age, sex, education, live alone/do not live alone), life style- smoking and alcohol drinking, health condition/history-presence of chronic disease (diabetes, cancer, stroke, cardiovascular disease). Estimates of risk were derived from the coefficient in cox regression model and 95% CI were calculated. A p-value of less than or equal to 0.05 was used to determine significant association in statistical testing. SPSS 11.0 for Windows was used to run the analyses.

3.8. DEFINITION AND MEASUREMENT OF VARIABLES

Three set of measures were relevant to this study; Outcome variable which is the number of years until death occurs, explanatory variable which is the social activity involvement status and the potential covariates which are the other variables that can contribute to the relationship between the outcome variable and the explanatory variable. These set of covariates includes socio-demographic characteristics such as age, gender, education; Life style such as participants living alone or not, smoke or not, drink alcohol or not; health conditions/history such as presence of any chronic disease and cognitive status. These variables were collected at baseline screening interview in 1992/93 with the community screening interview document, CSI-D.

Outcome variable

The outcome variable is the number of years until death occurs called Survival time.

Survival Time (t) is calculated by subtracting the date of death or lost to follow-up from the date that the subject entered into the study (March, 1992) and it was recorded in years.

The interviewers continuously monitored this densely populated and close knit Idikan community all through at regular interval, daily visit to the site throughout the course of the

11

Ibadan-Indianapolis Dementia Research Project and recorded all death and also movement of the subject in and out of the community in order to be able to account for and involve in the analysis the censored and uncensored data by taking the records of traveled, missing, moved out and lost to follow up status. By the end of March, 2009, 27.2% of the study size was recorded as deceased while 72.8% were found living.

Censored data: A Survival time, S (t) is being censored only when they are lost to follow-up (traveled, moved out of idikan town or Untraceable) or did not experience death until the end of the study in March, 2009. The censorship status assigned 'zero' for 'censored' (those alive), and 'one' for 'failed' (those dead). That is, N (0) = 677 persons while N (1) = 1808 persons.

Explanatory Variables

The explanatory variables involved in this study are further classified into two; the exposure variable and the covariates

Exposure Variable

The only exposure variable in this study is the status of social activity involvement. It is the exposure variable because it is the only primary variable of interest.

Social activity involvement: It is a baseline information collected during the screening interview in 1992/93, the subject were asked to indicate their involvement in some set of activities and the participant's relative who has knowledge of his/her daily functioning was also interviewed for confirmation especially where the subject has some form of disabilities or health conditions. The social activity index assigns 'one' point for 'not being socially active' and 'zero' for 'being socially active'.

'Not socially active' group refers to those who cannot carry out any waking day activities such as bathing, eating and resting, are assisted by others including those who are only productively active: that is, active at home during waking day. Such as takes care of self and

12

others including provision of cares for relatives, cook meal, do laundry or generally look after someone, house work including maintenance of home and possessions, running errands, paid work, receive/see visitors, friends/family seek advice from, gardening, sewing, knitting, dancing at home, wood/metal working, watching television/movies at home.

'Socially active' group refers to those who perform consumptive activities: that is, extensively active at home and outside home during waking day. Such as meet/visit friends and family, go to restaurant/centers/senior center /social club, go to play music/dancing outside / watch TV/movies, listen to music/records/tapes, watch/participate in sports, play games/puzzle/cards, fishing, attend parties/ Church/Mosques, neighborhood/local politics, attending adult education courses and other social interaction.

In this study, 374 persons were found to be 'not socially active' while 2111 persons were found to be 'socially active'.

Potential Confounders/Covariates /Extraneous variables

They are referred to as the potential confounders or covariates or extraneous variables because they are not the primary study and they can likely influence the relationship between the exposure (primary study) and outcome variable.

Socio-demographic variables

Sex: It is a dichotomous variable indicating whether a participant is a male or a female.

The sex index assigns 'one' for being a 'female' and 'zero' for being a 'male'. In this study, N (Female) = 1617 while N (Males) = 868.

Age: Age is a continuous data made categorical. Date of birth for the Yorubas were estimated from a table of historical land marks well-known to the population, a well-tested, long-standing practice in Nigeria for assessing ages of adults (Ogunniyi and Osuntokun, 1993). Age index was defined as the age as at the time of interview and was calculated by subtracting the date of birth from the date of interview (1992/93), assigning 'zero' point to

13

the age category 65-74 years, 'one' point to the age category 75-84 years and 'two' point to the age category 85years and above. That is, 65-74=0; 75-84=1; =85=2. n= 1792 persons were aged 65 to 74 years, n= 470 persons were aged 75 to 84 years while n= 223 persons were aged 85years and above.

Education: It is a dichotomous variable indicating whether or not a participant ever attends school. The education index assigns 'one' point for 'no' as a response to can read and write? And 'zero' point for 'yes' as a response to can read and write? n=377 persons reported can read and write while n= 2108 persons reported cannot read and write.

Life style variable

The life style variables collected were smoking and alcohol drinking history, Co-habiting status. Smoking and alcohol history were self-reported on the screening interview questionnaire

Smoking status: It is a dichotomous variable indicating whether or not a participant does smoke. The smoking index assigns 'one' point for smoking and zero point for not smoking. n=603 persons do smoke while n= 1883 persons do not smoke.

Alcohol status: It is a dichotomous variable indicating whether or not a participant does drink alcohol. The alcohol index assigns 'one' point for drinking alcohol and 'zero' point for not drinking alcohol. n= 614 persons drink alcohol while n=1871 persons do not drink alcohol.

House Composition: It is a dichotomous variable indicating the living arrangement of the participant, whether he/she lives alone or lives with others (his or her partner(s) alone or family no spouse, family with partners or others). This variable was self-reported on the screening interview with interviewer's observation. The house composition index assigns one point for living alone and zero point for living with others. n= 31 persons lives alone while n= 2454 persons lives with others.

14

Health Status Variables: The health status variables collected were presence of chronic disease and cognitive functioning status.

Presence of Chronic Disease: It is a dichotomous variable indicating whether or not the participant has the presence of at least one chronic disease such as hypertension, stroke, diabetes, cancer. This comorbidity index was calculated by scoring one for the presence of any of the chronic diseases mentioned earlier. n= 214 persons has the presence of at least one chronic disease while n= 2271 persons do not have any.

Those responding "don't know" for any particular condition were considered free of the condition.

Cognitive functioning: It is a continuous variable made categorical. All subjects were administered a Yoruba translated version of abbreviated mental test score to single out subject with cognitive impairment. Based on the cognitive scores reported by the CSI-D, subjects were classified into three groups namely: > 29.5 = Good performance, > 28.5 and ≤ 29.5 = Intermediate, ≤ 28.5 = Poor performance; with a low, medium, or high probability of being demented respectively (Hall et al., 1993; Hall et al., 1996). The cognitive functioning index therefore assigned 'zero' for good, 'one' for intermediate and 'two' for poor performance respectively. n= 1053 persons has good performance, n= 511 persons has intermediate performance and n= 921 persons has poor performance.

3.9. Exclusive & Inclusive criteria: Any person with missing data on any of the variables in consideration was excluded from the study. Therefore, the original 2,494 interviewed at the baseline survey reduced to a total studied population sample size of N=2,485 because of nine persons that were completely excluded from the final analysis due to missing information. No percentage untraceable or missing information on the variables that were involved in the final study due to the exclusion criteria.

15

3.10.0. SURVIVAL ANALYSIS

It is a collection of statistical procedures for data analysis for which the 'outcome variable of interest' is 'time until an event occurs'.

3.10.1. Terminologies and Key Notations

Terminologies

Event: By event, we mean death, disease incidence, relapse from remission, recovery (e.g., return to work) or any designated experience of interest that may happen to an individual. It is typically referred to as a 'failure', because the kind of event of interest usually is death, disease incidence or some other negative individual experience. However, survival time may be 'time to return to work after an elective surgical procedures' in which case failure is a positive event. That is, failure = Event of interest. However, failure can be positive event, such as returning to school.

Time: By time, we mean years, months, weeks or days from the beginning of follow-up of an individual until an event occurs; alternatively time can refer to the age of an individual when an event occurs.

Survival Time: In a survival analysis, we usually refer to the time variable as survival time, because it gives the time that an individual has 'survived' over some follow-up period. That is, Survival Time = Time. It is the outcome variable.

Outcome Variable: It is the variable that is the focus of our attention, whose variation or occurrence we are seeking to understand.

Explanatory variable: This is a factor that may influence the size, or the occurrence of the outcome variable. It can be classified into two; the exposure variable and the extraneous variables.

Exposure variable: It is an explanatory variable but the primary factor of interest that may influence the size or the occurrence of the outcome variable.

16

Extraneous variable: It is an explanatory variable but also an extraneous variable for it is the one we are interested in adjusting or controlling for, because of either its confounding or interactive effect.

Confounder(s): These is/are variable(s) that can affect the outcome-exposure association/relationship which are said to confound the association of interest in which, failure to control for them can lead to confounding bias. That is, it can confound the effect of exposure on the outcome.

Interaction: When the relationship between the exposure and outcome variable is known to be different depending on the level of the other variable, then there is an interaction.

Competing Risk: This statistical problem occurs when more than one events is considered in the same analysis (e.g., death from any of several causes). To solve this problem, although more than one event may be considered in the same analysis, we will assume that only one event is of design interest.

Censoring: It is a key analytical problem in survival analysis. Censoring occurs when we have some information about individual survival time, but we don't know the exact survival time. We usually want to use the observed survival time to draw implications about the true survival time so that we don't lose information and thereby generate a misleading result.

There are generally three reasons why censoring may occur:

. Study ends- no event: A person does not experience the event before the study ends

. Lost: A person is lost to follow-up during the study period

. Withdraws: A person withdraws from the study because of death (if death is not the designated event of interest) or some other reason (e.g., adverse drug reaction)

If a person does not fail, that is, does not get the event during the study period, censorship is the only remaining possibility for that person's survival time.

Survival time can be indicated as censored or failed with '1' denoting 'failed' and '0'

17

denoting 'censored'.

Two types of censored data are Right censored data and Left censored data.

Right censored data: These occur when the person's survival time becomes incomplete (has been cut off) at the right side of the follow-up period, that is, occurring when the study ends or when the person is lost to follow-up or is withdrawn. Most survival time is R.C. The survival time for this study is R.C.

Left Censored data: These occur when the person's survival time becomes incomplete (has been cut off) to the left side of the follow-up period for that person. For instance, if we are following persons with HIV infection, we may start follow-up when a subject first tests positive for the HIV virus, but we may not know exactly the time of first exposure to the virus. Thus, the survival time is censored on the left side, because there is unknown follow-up time from the time of first exposure up to time of first up to time of first positive HIV test.

Survivor and Hazard Function

The two quantitative terms considered in any survival analysis are 'Survivor function' and 'Hazard functions'. Of these two functions, S (t) and H (t), the survivor function is more naturally appealing for analysis of survival date, simply because S (t) directly describes the survival experience of a study cohort. However, the hazard function is also of interest for the following reasons:

- It provides insight about conditional failure rates;

- It may be used to identify a specific model form, such as an exponential, a Weibull, or a lognormal curve that fits one's data.

- It is the vehicle by which mathematical modeling of survival data is carried out; that is the survival model is usually written in terms of the hazard function.

18

denoting 'censored'.

Two types of censored data are Right censored data and Left censored data.

Right censored data: These occur when the person's survival time becomes incomplete (has been cut off) at the right side of the follow-up period, that is, occurring when the study ends or when the person is lost to follow-up or is withdrawn. Most survival time is R.C. The survival time for this study is R.C.

Left Censored data: These occur when the person's survival time becomes incomplete (has been cut off) to the left side of the follow-up period for that person. For instance, if we are following persons with HIV infection, we may start follow-up when a subject first tests positive for the HIV virus, but we may not know exactly the time of first exposure to the virus. Thus, the survival time is censored on the left side, because there is unknown follow-up time from the time of first exposure up to time of first up to time of first positive HIV test.

Survivor and Hazard Function

The two quantitative terms considered in any survival analysis are 'Survivor function' and 'Hazard functions'. Of these two functions, S (t) and H (t), the survivor function is more naturally appealing for analysis of survival date, simply because S (t) directly describes the survival experience of a study cohort. However, the hazard function is also of interest for the following reasons:

- It provides insight about conditional failure rates;

- It may be used to identify a specific model form, such as an exponential, a Weibull, or a lognormal curve that fits one's data.

- It is the vehicle by which mathematical modeling of survival data is carried out; that is the survival model is usually written in terms of the hazard function.

18

Key Notations

T: The random variable 'person survival time' is denoted by capital T. Since T denotes time, its possible values include all nonnegative numbers. That is, T = Survival time, (T = 0).

t: Any specific value of interest for the random variable, capital T is denoted by a small letter 't'. For example, if we are interested in evaluating whether a person survives for more than five (5) years after undergoing cancer therapy, it then means that small 't' equals '5', capital T exceeds '5'. t = specific value for T. Survives > 5 years implies T > t = 5.

3.10.2. Goals of Survival Analyses

The basic goals of survival analyses are to:

- Estimate and interpret survivor and/or hazard functions from survival data

- Compare survivor and/or hazard functions.

- Assess the relationship of explanatory variables to survival time

3.10.3. Survivor Function and Hazard Function

Survivor Function

It is denoted by S (t). It gives the probability that a person survives longer than a specified time, t. Survival function is fundamental to a survival analysis, because obtaining survival probabilities for different values of 't' provides crucial summary information from survival data. Theoretically, as t ranges from '0' to '∞'; S (t) can be graphed as a smooth curve.

All survivor functions have the following characteristics:

1. They are non-increasing; that is, they headed downward as to increasing.

2. At time t = 0, S(t) = S(0) = 1; that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time zero (0) is one (1).

3. At time t = ∞, S(t) = S(∞) = 0; That is, theoretically, if the study period increased

19

without limit, eventually no one will survive, so the survivor curve must eventually fall to zero.

4. But in practice, when using actual data, we usually obtain graphs that are step functions. Because, the study period is never infinite in length, it is possible that not everyone studied gets the event. In practice, the estimated survivor function, denoted by a caret over the S as in Ŝ(t), thus does not go all the way down to zero at the end of the study.

Hazard Function

It is denoted by h (t), is given by the formula:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

That is, h (t) equals limit, as Δt approaches zero, of a probability statement about survival, divide by Δt, where Δt denotes a small interval of time. The hazard function h(t) gives the instantaneous potential for failing at time 't' per unit time, given survival up to time t. h(t) is also called **instantaneous potential or conditional failure rate.**

For a specified value of t, the hazard function h(t) has the following characteristics:

1. $H(t) = 0$; it is always nonnegative, that is, equal to or greater than zero;

2. H(t) has no upper bound.

Types of Hazard Functions

- Exponential Model: Hazard function is constant at any given value of survival time 't'

- Increasing Weibull Model: H(t) increases at any increase in survival time 't'

- Decreasing Weibull Model: Hazard function decreases at any increase in survival time 't'

- Lognormal Survival Model: Hazard function decreases at the early stage of survival but then decreases later as the survival time 't' increases

20

3.10.4. Differences and Relationship of Survivor functions and Hazard functions

Differences

Hazard function h (t) unlike survivor function s(t) is a rate and not a probability because it is a ratio of two quantities; $P(t = T < t + \Delta t \mid T = t)$ and Δt. H(t) is time dependent. The scale for this ratio is not 0 to 1 like survivor function but 0 and 8 and depends on whether time is measured in days, weeks, months or years, etc.

In contrast to a survivor function, the graph of h (t) does not have to start at 1 and go down to zero, but rather can start anywhere and go up and down in any direction over time.

Relationship of S (t) and H (t)

There is a clearly defined relationship between the two. In fact, if one knows the form of S (t), one can derive the corresponding h (t) and vice versa.

The relationship is expressed in general formulae as follows:

1. $S\,(t) = \exp\left[-\int_0^t h\,(u)\,du\right]$

2. $h\,(t) = -\left[\dfrac{\left(\frac{dS(t)}{dt}\right)}{S(t)}\right]$

The first of these formulae describes how the survivor functions S (t) can be written in terms of an integral involving the hazard function. The formula says that S (t) equals the exponential of the negative integral of the hazard function between integration limits of time 'zero' and 't'.

The second formula describes how the hazard function h (t) can be written in terms of a derivative involving the survivor function. This formula says that h (t) equals minus the derivatives of S (t) with respect to t divided by S (t).

The higher the hazard rate, the lower is the group's probability of surviving.

21

3.10.5. Descriptive Measures of Survival Experience

The descriptive measures of survival experience are the Average Survival Time and the Average Hazard Rates. These descriptive measures provide the overall comparison of groups. Average Survival Time denoted by $\bar{T}$ is simply calculated by summing all the observed survival time $t_{(i)}$ in a group (ignoring whether a survival time is censored or failed) and then finding the average.

That is, $\bar{T} = \sum_{i=1}^{n} t_{(i)} / n$.

While the Average Hazard Rates denoted by $\bar{h}$ is defined by dividing the total number of failures by the sum of the observed survival times.

That is, $\bar{h}$ = no. of failures/ $\sum_{i=1}^{n} t_{(i)}$.

The hazard rate indicated failure potential rather than survival probability. Thus the higher the hazard rate, the lower is the probability of surviving.

3.10.6. Kaplan-Meier (KM) Method

Descriptive measures for survival experience cannot compare groups at different points in time of follow-up. Such comparison is provided by a graph of survivor curves generated by estimated survival probabilities at a given time. The method used to get these curves is called Kaplan- Meier (KM) method. It can be used to compute the survival probability at a given time using all the survival information collected on a censored person up to the time of censorship rather than simply throw away all the information on a censored person. Kaplan-Meier Method estimates survivor curves that are actually step functions and allows comparison of two groups at different points in time of follow-up. From the Kaplan-Meier graph, one can obtain the median survival time for a group graphically by proceeding horizontally from the 0.5 point on the Y-axis until survivor curve is reached, and then proceeding vertically downward until the X-axis is crossed at the median survival time.

22

Below is the basic data layout upon which Kaplan-Meier survival probabilities at a given time and curves are derived.

Ordered Data Layout:

Ordered Failure Times, $t_{(j)}$	no. of Failures, $m_{(j)}$	no. censored, $q_{(j)}$	Risk set, $n_{(j)}$	$\hat{S}(t_{(j)})$
$t_{(j)} = 0$	$m_{(j)} = 0$	$q_{(0)}$	$n_{(0)}$	$\hat{S}(t_{(0)})$
$t_{(1)}$	$m_{(1)}$	$q_{(1)}$	$n_{(1)}$	$\hat{S}(t_{(1)})$
$t_{(2)}$	$m_{(2)}$	$q_{(2)}$	$n_{(2)}$	$\hat{S}(t_{(2)})$
.				
.				
.				
$t_{(k)}$	$m_{(k)}$	$q_{(k)}$	$n_{(k)}$	$\hat{S}(t_{(k)})$

The first column in the table gives ordered survival times $t_{(j)}$ from the smallest to largest.

Each table begins with a survival time of zero, even though no subjects actually failed at the start of follow-up. The reason for the zero is to allow for the possibility that some subjects might have been censored before the earliest failure time.

The second column gives frequency counts of failure denoted by $m_{(j)}$ at each distinct failure time. The third column gives frequency counts, denoted by $q_{(j)}$, of those persons censored in the time interval starting with failure time $t_{(j)}$ up to but not including the next failure time denoted by $t_{(j+1)}$.

The fourth column gives the **risk set**, which denotes the collection of individuals who have survived at least to time $t_{(j)}$ or gives the number of subjects at the start of the interval. It is also denoted by $n_{(j)}$.

The last column denoted by $\hat{S}(t_{(j)})$ contains the survival probability estimates. The general formula for a KM survival probability at failure time $t_{(j)}$ is shown below:

23

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) \times Pr\,(T > t_{(j)} | \, T > t_{(j)}).$$

This formula gives the probability of surviving past the previous failure time $t_{(j-1)}$, multiplied by the conditional probability of surviving past time $t_{(j)}$, given survival to at least $t_{(j)}$.

The above KM formula can be expressed as a product limit if we substitute for the survival probability $\hat{S}(t_{(j-1)})$, the product of all fractions that estimate the conditional probabilities for failure times $t_{(j-1)}$ and earlier.

That is, $\hat{S}(t_{(j)}) = \prod_{i=1}^{j} Pr[T > t_{(i)} \,|\, T > t_{(i)}]$ (product limit formula)

$$= \hat{S}(t_{(j-1)}) \times Pr\,(T > t_{(j)} | \, T > t_{(j)})$$

Mathematical Proof:

Using the basic rules of probability, the KM formula can be described as follows;

Pr (A and B) = Pr (A) x Pr (B|A) always (rule of probability)

If we let A be the event that a subject survives to at least $t_{(j)}$, that is $A = $ " $T \geq t_{(j)}$ "

and we let B be the event that a subject survives past time $t_{(j)}$, that is, $B = $ ' $T > t_{(j)}$ '

Then the joint event A and B = B

It follows that, Pr (A and B) = Pr (B) = $S\,(t_{(j)})$

No failure during $t_{(j-1)} < T < t_{(j)}$

Therefore, Pr (A) = Pr (T > $t_{(j-1)}$) = $S\,(t_{(j-1)})$

Pr (B/A) = Pr (T > $t_{(j)}$ | T $\geq$ $t_{(j)}$)

24

Thus, from Pr (A and B) formula,

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) \times Pr\ (T > t_{(j)}|\ T > t_{(j)})$$

The KM survival probabilities and curves can be easily obtained from the most computer packages that perform survival analysis such as SPSS, SAS and the rest. All the user needs to do is provide a KM computer program with the basic data layout and then provide appropriate commands to obtain plots.

3.10.7.   Log-Rank Test for Two Groups and Several Groups

Log-Rank Test for Two Groups

It is the most popular testing method use to evaluate whether or not Kaplan-Meier (KM) survivor curves for two groups are statistically significant. When we state that two KM curves are "statistically equivalent", we mean that, based on a testing procedure that compares the two curves in some "overall sense" we do not have evidence to indicate that the true (population) survival curves are different.

The log-rank test is a large-sample chi-square test that uses as its criterion a statistics that provides an overall comparison of the KM curves being compared. This (log-rank) statistics, like many other statistics used in other kinds of chi-square tests, makes use of observed versus expected cell counts over categories of outcomes. The categories for the log-rank statistics are defined by each of the ordered failure times for the entire set of data being analyzed.

For each ordered failure time, $t_{(j)}$ in an entire set of data, the numbers of subjects failing at that time is denoted by $m_{(ij)}$, separately by group (i) as $m_{(1j)}$ and $m_{(2j)}$. Followed by the

25

numbers of subjects in the risk set at that time denoted by $n_{(ij)}$, separately by group (i) as $n_{(1j)}$ and $n_{(2j)}$.

The formula for the expected cell counts is given below:

For group 1,

$$e_{ij} = \left[\frac{n_{1j}}{n_{1j} + n_{2j}}\right] \times (m_{1j} + m_{2j})$$

$$\uparrow \qquad\qquad\qquad \uparrow$$

Proportion in the risk set   *no. of failures over both groups*

This formula computes the expected number at time j (i.e., e_{ij}) as the proportion of the total subjects in both groups who are at risk at time j, that is ($\frac{n_{1j}}{n_{1j} + n_{2j}}$) multiplied by the total number of failures at that time over both groups (i.e. $m_{1j} + m_{2j}$).

For group 2, e_{2j} is computed similarly as $e_{2j} = \left[\frac{n_{2j}}{n_{1j} + n_{2j}}\right] \times (m_{1j} + m_{2j})$

The log-rank test statistics is formed using the sum of the observed minus expected counts over all failure times for one of the two groups.

That is, $O_i - E_i = \sum_{j=1}^{n} (m_{ij} - e_{ij})$, i = 1, 2

The log-rank statistics is compared by dividing the square of the summed observed minus computed by dividing the square of the summed observed minus expected score for one of the groups by the estimated variance of the summed observed minus expected score.

That is, log-rank statistics $= \left[\frac{(O_i - E_i)^2}{Var(C_i - E_i)}\right]$

26

$$Var\left(O_i - E_i\right) = \sum_{i=1}^{n}\left(\frac{n_{1j}\,n_{2j}\,(m_{1j} + m_{2j})\,(n_{2j} + n_{2j} - m_{1j} - m_{2j})}{(n_{2j} - n_{2j})^2\,n_{1j} + n_{2j} - 1}\right), i = 1, 2$$

H_o : No difference between survival curves.

This variance formula involves the number in the risk set in each group (n_{ij}) and the number of failures in each group (m_{ij}) at time j. The summation is overall distinct failure times.

The null hypothesis being tested is that there is no overall difference between the two survival curves. Under this null hypothesis, the log-rank statistics is approximately chi-square with one degree of freedom. Thus, a p-value for the log-rank test is determined from tables of the chi-square distribution.

Several computer programs are available for calculating the log-rank statistics without the rigorous attempt of using the formula such as SPIDA, SPSS, SAS and the rest. During the computer printout interpretation, the p-value is an indicator of whether the null-hypothesis should be rejected or we fail to reject.

An approximation to the log-rank statistics can be calculated using observed and expected values for each group without having to complete the variance formula. The approximate formula is of the classic chi-square form that sums over each group being compared the square of the observed minus expected value divided by the expected value.

Approximate formula: $\chi^2 = \sum_{i=1}^{no\ of\ groups}\frac{(O_i - E_i)^2}{E_i}$

Log-Rank Test for Several Groups

The log-rank test can be used to compare three or more survival curves. The null hypothesis for this more general situation is that all survival curves are the same.

Formula Log-Rank Test for Several Groups:

27

For i = 1, 2. . . G and $j = 1, 2 \ldots k$, where G = no. of groups and k = no. of distinct failure times,

n_{ij} = no. at risk in ith group at jth ordered failure time

m_{ij} = observed no. of failures in ith group at jth ordered failure time

e_{ij} = expected no of failures in ith group at jth ordered failure time

$$= \left[\frac{n_{ij}}{n_{1j} + n_{2j}} \right] (m_{1j} + m_{2j})$$

$$n_i = \sum_{i=1}^{G} n_{ij}$$

$$m_i = \sum_{i=1}^{G} m_{ij}$$

$$O_i - E_i = \sum_{i=1}^{k} (m_{ij} - e_{ij})$$

$$\widetilde{Var}(O_i - E_i) = \sum_{i=1}^{k} \frac{n_{ij}(n_i - n_{ij}) m_i (n_i - m_i)}{n_i^2 (n_i - 1)}$$

$$\widetilde{Cov}(O_i - E_i, O_l - E_l) = \sum_{i=1}^{k} \frac{-n_{ij} n_{lj} m_i (n_i - m_i)}{n_i^2 (n_i - 1)}$$

$$d = (O_1 - E_1, O_2 - E_2 \ldots, O_{G-1} - E_{G-1})$$

$v = v_{ii} = \widetilde{Var}(O_i - E_i)$ and $v_{il} = \widetilde{Cov}(O_i - E_i, O_l - E_l)$ for i = 1, 2. . . G – 1;

l = 1, 2 . . . , G – 1.

Then the log-rank statistics is given by the matrix product formula

Log-rank statistics = $d' v^{-1} d$, which has approximately a chi-square distribution with G – 1 degree of freedom under the null hypothesis that all G groups have a common survival curve.

3.10.8. Cox Proportional Hazard Model, Formula and Properties

Cox proportional hazard model

Cox proportional hazard regression also known as **Cox regression** is the most commonly used approach to the regression of survival data. The model is a popular mathematical model used for analyzing survival data with the basic question of interest which concerns: comparing of survival experience of groups adjusting for or evaluating the possible confounding and/or interaction effect of covariates on the outcome variables.

The formula for Cox proportional hazard model

The Cox proportional hazards (PH) model is usually written in terms of the hazard model formula. This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables.

$$h(t, X) = h_0(t) \; e^{\sum_{i=1}^{p} \beta_i X_i}$$

$X = (X_1, X_2, \ldots, X_p)$ explanatory/predictor variable.

The Cox model formula says that the hazard at time t is the product of two quantities. The first of these, $h_0(t)$, is called the baseline hazard function. The second quantity is the exponential expression e to the linear sum of $\beta_i X_i$, where the sum is over the p explanatory X variable. An important feature of this formula, which concerns the proportional hazards (PH) assumption, is that the baseline hazards is a function of t, but does not involve the X's. In contrast, the exponential expression shown here, involves the X's, but does not involve t. The X's here are called time-independent X's. It is possible nevertheless, to consider X's which do not involve t. Such X's are called time-dependent variable. If time-dependent

29

variables are considered, the Cox model form may still be used, but such a model no longer satisfies the PH assumptions and is called the extended Cox model.

Time - Independent Variables

It is defined to be any variable whose values for a given individual do not change over time. However, a variable values may change over time, but for purposes of the analysis, it is assumed not to change once it is measured, so that only one value per individual is used and also if the effect of such variables on survival risk depends essentially on the value at only one measurement.

Properties of Cox model formula

1. The Cox model formula has the property that if all the X's area equal to zero, or when no X's are in the model, the formula reduces to the baseline hazard function. That is, the exponential part of the formula becomes e to the zero, which is one. This property of the Cox model is the reason why $h_0(t)$ is called baseline function.

That is, $h(t, \mathbf{X}) = h_0(t) \, e^{\sum_{i=1}^{p} \beta_i X_i}$

If $X_1 = X_2 \cdots = X_k = 0$

$= h_0(t) \, e^0$

$= h_0(t)$

$=$ Baseline hazard.

Or when no X's are in the model, thus $h_0(t)$ may be considered as a starting or "baseline" version of the hazard function, prior to considering any of the X's.

That is, $h(t, \mathbf{X}) = h_0(t)$

2. Another important property of the Cox model is that the baseline hazard, $h_0(t)$, is an unspecified function. It is this property that makes the Cox model a nonparametric model. In contrast, a parametric model is one whose functional form is completely specified, except for the values of the unknown parameters.

3.10.9. Why the Cox PH Model is Popular and its ML Estimation

Why the Cox PH model is popular

1. A key reason for the popularity of the Cox model is that it is nonparametric. Even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratio of interest and adjusted curves can be obtained for a wide variety of data situations. Another way of saying this is that the Cox PH model is a "robust" model, because the results from using the Cox model will closely approximate the results for the correct parametric model. When in doubt, as is typically the case, the Cox model is a "safe" choice of model.

2. In addition to the general robustness of the Cox mode, the specific form of the model is attractive for several reasons. The specific form of the Cox model gives the hazard function as a product of a baseline hazard involving t and an exponential expression involving the X's without t. The exponential part of this product is appealing because it ensures that the fitted model will always give estimated hazards that are non-negative. By definition, the values of any hazard function must range between zero and plus infinity, that is, a hazard is always non-negative.

3. Another appealing property of the Cox model is that, even though the baseline hazard part of the model is unspecified, it is still possible to estimate the β's in the exponential part of the model. That is, the primary information desired from a survival analysis namely the hazard ratio (HR) and survival curve can be obtained

31

using a minimum of assumptions without having to estimate the baseline hazard function $h_0(t)$.

4. One last point about the popularity of the Cox model is that, it is preferred over the logistic model when survival time information is available and there is censoring. That is, the Cox model uses more information- the survival times- than the logistic model, which considers a $(0,1)$ outcomes and ignores survival times and censoring.

ML Estimation of the Cox PH model

The method of estimation used to obtain the coefficients for Cox regression model is called, the maximum likelihood estimation. The estimates of the Cox model parameters $\beta's$ in the general formula shown as follow, $h(t, X) = h_0(t) e^{\sum_{i=1}^{f} \beta_i z_i}$ are called maximum likelihood (ML) estimates and are denoted as $\widehat{\beta_i}$. The ML estimates of the Cox model parameters are derived by maximizing a likelihood function usually denoted as L. The likelihood function is a mathematical expression which describes the joint probability of obtaining the data actually observed on the subjects in the study as a function of the unknown parameters (the $\beta's$) in the model being considered. L is sometimes written notationally as L $(\beta's)$ where β denotes the collection of unknown parameters.

The mathematical expression for L for the Cox model is quite complicated; moreover, in practice, the formula for L is built into the computer program used for survival analysis, so, there is no need to see it in order to obtain the ML estimates.

It is usually called a "partial" likelihood function rather than a (complete) likelihood function. It is called "partial" likelihood because the likelihood formula considers probabilities only for those subjects who fail, and does not explicitly consider probabilities for those subjects who

32

are censored. Thus the likelihood for the Cox model does not consider probabilities for all subjects, and so it is called a "a partial" likelihood.

In particular, the partial likelihood can be written as the product of several likelihoods, one for each of, say k failure times. Thus, at the jth failure time, L_j denote the likelihood of failing at this time, given survival up to this time. Note that the set of individuals at the jth failure time is called the "risk set" $R(t_{(j)})$ and this set will change- actually get smaller in size as the failure time increases.

$L = L_1 \times L_2 \times L_3 \times \ldots .L_k = \prod_{j=1}^{k} L_j$ where L_j = portion of L for the jth failure time given the risk set $R(t_{(j)})$.

Thus, although the partial likelihood focuses on subjects who fail, survival time information prior to censorship is used for those subjects who are censored. That is, a person who is censored after the jth failure time is part of the risk set used to compute L_j even though this person is censored later.

Steps for obtaining ML estimates:

1. The likelihood function is formed for a given model

2. Maximize the natural log of L. The maximization process is carried out by taking partial derivatives of L with respect to each parameters in the model,

3. Then, solving a system of equations as shown below:
   $\frac{\partial L}{\partial \beta_i} = 0, i = 1, \ldots, p$ (number of parameters)
   This solution is carried out using iteration. That is, the solution is obtained in a stepwise manner, which starts with

   (i)   A guessed value for the solution

   (ii) Then successively modifies the guessed value until a solution is finally obtained.

33

The ML estimates are used in obtaining the three statistical objectives typically considered in survival analysis. The major test statistics used with ML estimates are the **Wald statistics** and the **Likelihood ratio or LR statistics** which makes use of the log likelihood value.

3.10.10.   Computing the Hazard Ratio

In general, a hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the **X's**.

That is, $\widehat{HR} = \dfrac{\hat{h}\,(t,X^*)}{\hat{h}\,(t,X)}$

Where $X^* = X_1{}^*, X_2{}^*, \ldots, X_p{}^*$

And $X = (X_1, X_2, \ldots, X_p)$ denote the set of **X's** for two individuals

It is easier to interpret an HR that exceeds the null value of 1 than an HR that is less than 1.

That is, $\hat{h}\,(t, X^*) \geq \hat{h}\,(t, X)$. The **X's** are typically coded so that the group with the smaller hazard corresponds to **X**.

An expression is obtained for the HR formula in terms of the regression coefficients by substituting the Cox model formula into the numerator and denominator of the hazard ratio expression. This substitution is shown below.

That is, $\widehat{HR} = \dfrac{\hat{h}\,(t,X^*)}{\hat{h}\,(t,X)} = \dfrac{h_0\,(t)\,e^{\sum_{i=1}^{p}\hat{\beta}_i x^*_i}}{h_0\,(t)\,e^{\sum_{i=1}^{p}\hat{\beta}_i x_i}}$

$$= e^{\sum_{i=1}^{p}\hat{\beta}_i (X^*_i - X_i)}$$

$$= \exp\left[\sum_{i=1}^{p}\hat{\beta}_i (X^*_i - X_i)\right]$$

34

General Rules:

1. The hazard ratio for the effect of a (0, 1) exposure variable which adjusts for other variables is obtained by the exponentiation of the estimated coefficient of the exposure variable. This rule has the proviso that the model does not contain any product terms involving exposure.

   That is, if X_1 is a (0, 1) exposure variable,

   then $\widehat{HR} = e^{\widehat{\beta_1}}$ (= effect of exposure adjusted for other X's) provided no other X's are product terms involving exposure.

2. The hazard ratio for the effect of a (0,1) exposure variable in a model which contains product terms involving this exposure with others X's can be written as follows:

   $$\widehat{HR} = \exp\left[\widehat{\beta} + \sum \widehat{\delta_j} W_j\right]$$

   Where

   $\widehat{\beta}$ = coefficient of exposure variable, E

   $\widehat{\delta_j}$ = coefficient of product terms, E x W_j

   ($\widehat{HR}$ does not contain coefficients of non-products term)

3.10.11. Adjusted Survival Curves using the Cox Proportional Hazard (PH) Model

If no model is used to fit survival data, a survival curve can be estimated using a Kaplan-Meier method. Such KM curves are plotted as step functions. When a Cox model is used to fit survival data, survival curves can be obtain that adjust for the explanatory variables used as predictors. These are called adjusted survival curve, and, like Kaplan-Meier (KM) curves, these are also plotted as step functions. It is mathematically different from KM curves. The KM curves do not adjust for covariates and therefore, are not computed using result from a

35

fitted Cox PH model. The adjusted survival curve are computed using results from a fitted Cox PH model. The hazard function formula for the Cox PH model can be converted to a corresponding survival function formula for the Cox model as shown below:

Cox model hazard function: $h(t, X) = \widehat{h_0}(t) \, e^{\sum_{i=1}^{p} \beta_i x_i}$

Cox model survival function: $S(t, X) = S_0(t)^{e^{\sum \beta_i x_i}}$

This survival function formula is the basis for determining adjusted survival curves.

The expression for the estimated survival function can be written as

$$\hat{S}(t, X) = [\widehat{S_0}(t)]^{e^{\sum \hat{\beta}_i x_i}}$$

$\widehat{S_0}(t)$ and $\hat{\beta}_i$ are provided by the computer program that fits the Cox model. The X's, however, must first be specified by the investigation before the computer program can compute the estimated survival curve.

Typically, when computing adjusted survival curves, the values chosen for a covariate being adjusted is an average value like an arithmetic mean or a median. In fact, most computer programs for the Cox model automatically use the mean value over all subjects for each covariate being adjusted.

General formula for adjusted survival curves comparing two groups is shown below:

Note that we are assuming that the exposure variable is variable X_1, whose estimated coefficient is $\hat{\beta}_i$ and the value of X_1 is 1 for exposed and 0 for unexposed subjects.

Where Exposed subjects $X_1 = 1$

$$S(t, X_1) = [\widehat{S_0}(t)]^{e^{\hat{\beta}_1(1) + \sum_{i=1} \hat{\beta}_i X_i}}$$

Unexposed subjects $= X_0 = 0$

$$\check{S}(t, X_0) = [\widehat{S_0}(t)]^{\text{v } \widehat{\beta_2}(0) + \sum_{i=2} \widehat{\beta_i} \overline{X_i}}$$

General formula for adjusted survival curve for all covariates in the mode

$$\hat{S}(t, X_1) = [\widehat{S_0}(t)] \cdot^{\sum \widehat{\beta_i} \overline{X_i}}$$

This formula will give a single adjusted survival curve rather than different curves for each exposure group.

3.10.12. Meaning and Properties of Proportional Hazard (PH) Assumption

The PH assumption requires, that the ratio of hazard comparing different exposure group remains constant over time. Or equivalently, that the hazard for one individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time.

Re-considering the final expression for the hazard ratio,

$$HR = \frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = \exp\left[\sum_{i=1}^{p} \widehat{\beta_i}(X^*_i - X_i)\right]$$

Thus once the model is fitted and the values for X^* and X are specified, the value of the exponential expression for the estimated hazard ratio is a constant, which does not depend on time. If we denote this constant by $\hat{\theta}$,

That is, let $\hat{\theta} = \exp\left[\sum_{i=1}^{p} \widehat{\beta_i}(X^*_i - X_i)\right]$

Then we can write the HR as

37

$$HR = \frac{\hat{h}(t,X^*)}{\hat{h}(t,X)} = \hat{\theta}$$

This is a mathematical expression which states the proportional hazards assumption. Another way to write the proportional hazards assumption mathematically expresses the hazard function for individual X, as shown below;

That is $\hat{h}(t,X^*) = \hat{\theta}\ \hat{h}(t, X)$

This expression says that the hazard function for one individual is proportional to the hazard function for another individual, where the proportionality constant is $\hat{\theta}$ which does not depend on time.

In general, if the hazards cross, then the PH assumption cannot be met, in that, if the hazard ratio varies over time, then a Cox PH model is not appropriate. The P (PH) information allows one to evaluate the proportional hazards (PH) assumption. The p-value can be derived from a standard normal statistics computed from the Cox proportional hazard model output. A non-significant (i.e. large) p-value, say greater than 0.05 indicates that the PH assumption I satisfied, whereas a small p-value, say < 0.05 indicates that the variable being tested does not satisfy this assumption. The P (PH) approach for evaluating whether the PH assumption is satisfied is only one of three approaches.

3.10.13.   Interpreting the result of Cox PH Analysis

For any Cox proportional hazard model, the basic question is assessing the effect of exposure variable on outcome variable adjusting for the possible confounding and interaction effect. In statistical analyses, result are presented by reporting a confidence interval (C.I) which give a range of likely values for the difference in the population, and a p-value which addresses

38

whether the observed difference in the sample could arise because of chance alone, if there were no difference in the population.

In this study, where the question is to assess the relationship or association (the difference in effect) between the exposure variable "status of involvement in social activity" and outcome variable " death" , there are three statistical objectives typically considered:

1. Test for significance of effect- Test of hypothesis
2. Point estimate of effect
3. Confidence interval for effect

Before interpreting any computer printout of a Cox regression analysis, a fitted Cox PH model needed to be identify by putting into consideration the potential of having interaction effect and confounding effect of the covariates.

In other to interpret the computer printout of a Cox regression analysis, it is necessary to identify the following:

1. The variables that are included in the model
2. Regression coefficients corresponding to each variable in the model
3. Standard errors of the regression coefficients
4. The p-values for testing the significance of each coefficient where not given can be calculated by dividing the coefficient of the independent variables by its standard error and whatever it gives, it is assume that the quantity is approximately a standard normal or Z variable. This Z statistics is known as **Wald statistics**, which is one of the two test statistics typically used with ML estimates. The other test statistics called **Likelihood ratio or LR statistics**, makes use of the log likelihood value.

whether the observed difference in the sample could arise because of chance alone, if there were no difference in the population.

In this study, where the question is to assess the relationship or association (the difference in effect) between the exposure variable "status of involvement in social activity" and outcome variable " death" , there are three statistical objectives typically considered:

1. Test for significance of effect- Test of hypothesis

2. Point estimate of effect

3. Confidence interval for effect

Before interpreting any computer printout of a Cox regression analysis, a fitted Cox PH model needed to be identify by putting into consideration the potential of having interaction effect and confounding effect of the covariates.

In other to interpret the computer printout of a Cox regression analysis, it is necessary to identify the following:

1. The variables that are included in the model

2. Regression coefficients corresponding to each variable in the model

3. Standard errors of the regression coefficients

4. The p-values for testing the significance of each coefficient where not given can be calculated by dividing the coefficient of the independent variables by its standard error and whatever it gives, it is assume that the quantity is approximately a standard normal or Z variable. This Z statistics is known as **Wald statistics**, which is one of the two test statistics typically used with ML estimates. The other test statistics called **Likelihood ratio or LR statistics**, makes use of the log likelihood value.

39

A p-value is obtained for the coefficient of the independent variables. This p-value is an indicator of whether there is or there is no significant group effect or confounding effect or interaction effect.

5. A point estimate of effect is provided by the estimated hazard ratio (HR) value for the effect of the variable. If not given, it is given by taken the exponential of the coefficient of the explanatory variable.

6. To describe the confidence interval (C.I) for the effect of explanatory/independent variable. It is the C.I for the HR value which is given in the computer printout. If not given, it can be calculated as follows: For instance, to compute a 95% C.I for the regression coefficient of the explanatory variable, the large sample formula is the regression coefficient $\beta_i \pm 1.96 \times S.E$ where 1.96 is the 97.5 percentile of the standard normal or Z distribution. Followed by taken the exponential of the two limits obtained for the C.I for the regression coefficient of the explanatory variable.

3.10.14. Interpreting Statistical Analysis Result

Testing Hypothesis

Testing hypothesis is a process of disproving hypotheses. Statistical methods formulate this idea by looking for evidence against a very specific form of hypothesis, called a null hypothesis: that there is no difference between groups or no association between variables. Relevant data are then collected and assessed for their consistency with the null hypothesis. Links between exposure and outcomes, or between treatments and outcomes are assessed by examining the strength of the evidence against the null hypothesis, as measured by p-value.

In some circumstances, where the results hit one straight between eyes, statistical methods are not required in order to reject the null hypothesis. Unfortunately these situations are rare in medical research because there is rarely a one-to-one link between exposures and

40

outcomes; there is usually more inherent variability from person to person. The outcome can be unpredictable and is influenced by many other factors. Statistical methods are used to assess the strength of evidence against a null hypothesis, taking into account this person- to-person variability or things happening by chance. There comes the need for calculating test statistics and its corresponding p-value also known as a significant level.

General form of Confidence Interval (C.I) and Test Statistics

In all case, the C.I is constructed as the sample estimate (be it a mean, a difference between means or any of the other measures of exposure effect) plus or minus its standard error multiplied by the appropriate percentage point. Unless the sample size is small, this percentage point is based on the normal distribution (e.g. 1.96 for 95% C.Is).

95% C.I = Estimate – (1.96 X S.E) to Estimate + (1.96 X S.E).

The test statistics is simply the sample estimate divided by its standard error.

That is, Test statistics $= \dfrac{Estimate}{S.E}$

The test statistics measures by how many standard errors the estimate differs from the null of zero. It is used to derive a p-value. The standard error is inversely related to the sample size. Thus the larger the sample size, the smaller will be the standard error. Since the standard error determines the width of the C.I and the size of the test statistics, this also implies the following: For any particular size of difference between the two groups, the larger the sample size, the smaller will be the C.I and the larger the test statistics.

P- value is defined as the probabi liy of getting a difference at least as big as that observed if the null hypothesis is true. By convention, we usually use two-sided p-values; we include the possibility that the difference could have been of the same size but in the opposite direction.

41

The larger the test statistics, the smaller is the p-value. The interpretation of p-value is the same, no matter how they are derived.

Interpretation of P- Value

P- Value tells us the strength of the evidence against the null hypothesis that the true difference in the population is zero. The smaller the p-values, the stronger the evidence against the null-hypotheses. It has been a common practice to interpret a p-value by examining whether it is smaller that particular threshold value. In particular p-values less than 0.05 are often reported as 'statistically significant' and interpreted as being small enough to justify rejection of the null hypothesis. This is why hypothesis tests have been called significance tests. The 0.05 threshold is an arbitrary one that became commonly used in medical and psychological research, largely because p-value were determined by comparing the test statistics against tabulations of specific percentage points of distributions such as the z and t distributions.

In reporting the result of a study, it is recommended that the precise p-value should be reported together with the 95% C.I (Betty R.Kirkwood and Jonathan A.C Sterne, Essential Medical Statistics).

It should be acknowledged that the 95% C.I is based on the same arbitrary value as the 0.05 threshold; a z-value of 1.96 corresponds to a p-value of 0.05. This means that if p<0.05, then the 95% C.I will not contain the null value. However, the interpretation of a C.I should not focus on whether or not it contains the null value, but on the range and potential importance of the different values in the interval.

Since both C.I and p-values are derived from the size of the difference and it S.E, they are of course closely related. For example, if the 95% C.I does not contain the null value, then we

42

know that the p-value must be smaller than 0.05 and vice versa; if the 95% C.I does not include the null value, then the p-value will be greater than 0.05. Because the S.E decreases with increasing sample size, the width of the C.I and the size of the p-value are as dependent on the sample size as on the underlying population difference for a particular size of difference in the population, the larger the test statistics and the smaller the p-value.

In conclusion, in interpreting the results of the medical research, both the C.Is and p-values are very helpful.

43

CHAPTER 4

4.0. RESULT

Table 1 showed the frequency distribution of the elderly based on their background information. The total number of the studied population was 2, 485 with a male to female ratio of 1:2; 9% above 85 years and majority (72.1%) between the age of 65 – 74 years. The table showed that the literacy level is only 15% while only 1.2% lives alone. About a quarter drink alcohol while a quarter also smoke.

A little above a third had poor cognitive impairment while less than 10% had one kind of chronic disease. About 85% of the population was socially active.

44

Table 1: Frequency distribution of the baseline characteristics of the elderly in Idikan, 1992

Variable	Frequency	Percentage (%)
Age Group(years)		
65 – 74	1792	72.1
75 - 84	470	18.9
84+	223	9.0
Total	**2485**	**100**
Gender		
Male	868	34.9
Female	1617	65.1
Total	**2485**	**100**
Literacy Status		
Can read and write	377	15.2
Cannot read and write	2108	84.8
Total	**2485**	**100**
House Composition		
Lives alone	31	1.2
Do not live alone	2454	98.8
Total	**2485**	**100**
Drinks Alcohol		
Yes	614	24.7
No	1871	75.3
Total	**2485**	**100**
Smokes		
Yes	603	24.3
No	1882	75.7
Total	**2485**	**100**
Cognitive Status		
Good performance	1053	42.4
Intermediate performance	511	20.6
Poor performance	921	37.1
Total	**2485**	**100**
Presence of Chronic Disease(s)		
Yes	214	8.6
No	2271	91.4
Total	**2485**	**100**
Status of Participation in Social Activities		
Socially active	2111	84.9
Non active	374	15.1
Total	**2485**	**100**

Table 2 showed the relationship of baseline characteristics of the elderly in 1992 with their social activities status. The percentage socially active decreased with increase in age. For those less than 75years, it was 88% and then decreased to 70% for years higher. Also the female had higher proportion of those socially active than the male. The literacy status had no bearing with social activity. But there was a high significant statistical association between living arrangement, cognitive status, presence of chronic disease and social activity while consumption of alcohol and cigarette smoking were also significant at the 5% level.

Table 2: Relationship of the baseline characteristics of the elderly in Idikan (in 1992), with their status of 'involvement in social activities'

Variable	Socially Active	Not Socially Active	Column Total	P-value
Age Group(years)				
65 – 74	1577(88%)	215(12%)	1792(100%)	
75 - 84	378(80.4%)	92(19.6%)	470(100%)	
85+	156(70%)	67(30%)	223(100%)	
Total			2485(100%)	0.000
Gender				
Male	719(82.8%)	149(17.2%)	868(100%)	
Female	1392(86.1%)	225(13.9%)	1617(100%)	
Total			2485(100%)	0.031
Literacy Status				
Can read and write	321(85.1%)	56(14.9%)	377(100%)	
Cannot read and write	1790(84.9%)	318(15.1%)	2108(100%)	
Total			2485(100%)	0.908
House Composition				
Lives alone	18(58.1%)	13(41.9%)	2454(100%)	
Do not live alone	2093(85.3%)	361(14.7%)	31(100%)	
Total			2485(100%)	0.000
Drinks Alcohol				
Yes	501(81.6%)	1136(18.4%)	614(100%)	
No	1610(86.1%)	261(13.9%)	1871(100%)	
Total			2485(100%)	0.007
Smokes				
Yes	497(82.4%)	106(17.6%)	603(100%)	
No	1614(85.8%)	268(14.2%)	1882(100%)	
Total			2485(100%)	0.046
Cognitive Performance Status				
Good performance	907(86.1%)	146(13.9%)	1053(100%)	
Intermediate performance	457(89.4%)	54(10.6%)	511(100%)	
Poor performance	747(81.1%)	174(18.9%)	921(100%)	
Total			2485(100%)	0.000
Presence of Chronic Disease(s)				
Yes	150(70.1%)	64(29.9%)	214(100%)	
No	1961(86.3%)	310(13.7%)	2271(100%)	
Total			2485(100%)	0.000

47

Table 3 showed the mortality pattern of the elderly in Idikan with their background characteristics, 17 years after the baseline study in 1992. A higher proportion of those who were not socially active had died compare to those who were socially active. Also, a higher proportion of those who had poor cognitive status had died including those ever smoke and drink. It was also observed that literacy status had no bearing to death. A higher proportion of female are still alive as at study end.

Table 3: Mortality pattern of the elderly in Idikan, 17 years after the start of the study in 1992, according to their Baseline Characteristics.

Variable	Alive	Dead	Column Total	P-value
Age Group(years)				
65 – 74	605(33.8%)	1187(66.2%)	1792(100%)	
75 - 84	60(12.8%)	410(87.2%)	470(100%)	
84+	12(5.4%)	211(94.6%)	223(100%)	
Total			2485	0.000
Gender				
Male	167(19.2%)	701(80.8%)	868(100%)	
Female	510(31.5%)	1107(68.5%)	1617(100%)	
Total			2485	0.000
Literacy Status				
Can read and write	105(27.9%)	272(72.1%)	377(100%)	
Cannot read and write	572(27.1%)	1536(72.9%)	2108(100%)	
Total			2485	0.773
House Composition				
Lives alone	10(32.3%)	21(67.7%)	31(100%)	
Do not live alone	667(27.2%)	1787(72.8%)	2454(100%)	
Total			2485	0.528
Drinks Alcohol				
Yes	142(23.1%)	472(76.9%)	614(100%)	
No	1535(28.6%)	1336(71.4%)	1871(100%)	
Total			2485	0.008
Smokes				
Yes	125(20.7%)	478(79.3%)	603(100%)	
No	552(29.3%)	1330(70.7%)	1882(100%)	
Total			2485	0.000
Cognitive Status				
Good performance	325(30.9%)	728(69.1%)	1053(100%)	
Intermediate performance	139(27.2%)	372(72.8%)	511(100%)	
Poor performance	213(23.1%)	708(76.9%)	921(100%)	
Total			2485	0.001
Presence of Chronic Disease(s)				
Yes	30(14%)	184(86%)	214(100%)	
No	647(71.5)	1624(71.5%)	227(100%)	
Total			2485	0.000
Status of Participation in Social Activities				
Socially active	625(29.6%)	1486(70.4%)	2111(100%)	
Non active	52(13.9%)	322(86.1%)	374(100%)	
Total			2485	0.000

49

Table 4 showed the survival rate of the subjects over a period of 17 years (1992-2009) in Idikan. At the start of the study, since no one had gotten the event yet, the probability of surviving past time t equals zero (0) was exactly 0.984, approximately one (1).The probability of surviving past time t equals seventeen, at the end of the study was 0.265.

Table 4: The Elderly yearly survival rate in Idikan from 1992 – 2009

Year	Cum Survival	Standard error
0	0.984	0.003
1	0.944	0.005
2	0.874	0.007
3	0.784	0.008
4	0.686	0.009
5	0.612	0.010
6	0.551	0.010
7	0.507	0.010
8	0.474	0.010
9	0.442	0.010
10	0.415	0.010
11	0.385	0.010
12	0.360	0.010
13	0.336	0.009
14	0.318	0.009
15	0.295	0.009
16	0.280	0.009
17	0.265	0.009

50

Figure 1 is a Kaplan Meir Curve showing 17years survival pattern of the elderly in Idikan, based on their status of involvement in social activities at the beginning of the study in 1992. The graph showed that the survivor function for the 'socially active group' consistently lies above that of the 'not socially active' group and there was a widening gap between the two curves. This indicated that participation in social activity appeared effective against death at all points of follow-up and more effectively, during follow-up than it was early on. The curves showed that those who were not socially active presented with a poor survival than those who were socially active and it was statistically significant at 5% levels.

Table 5 showed the overall median survival time of the elderly in Idikan for the 17 years follow up. It also showed the median survival time for the two categories of the elderly based on their status of involvement in social activity. The overall median survival time was 8 years while the median survival time for socially active group and not socially active group were 9 years and 5 years respectively.

Figure 1: Kaplan Meir Curve showing the

Survival function of the different category of elderly in Idikan, Ibadan based on their status of participation in social activities
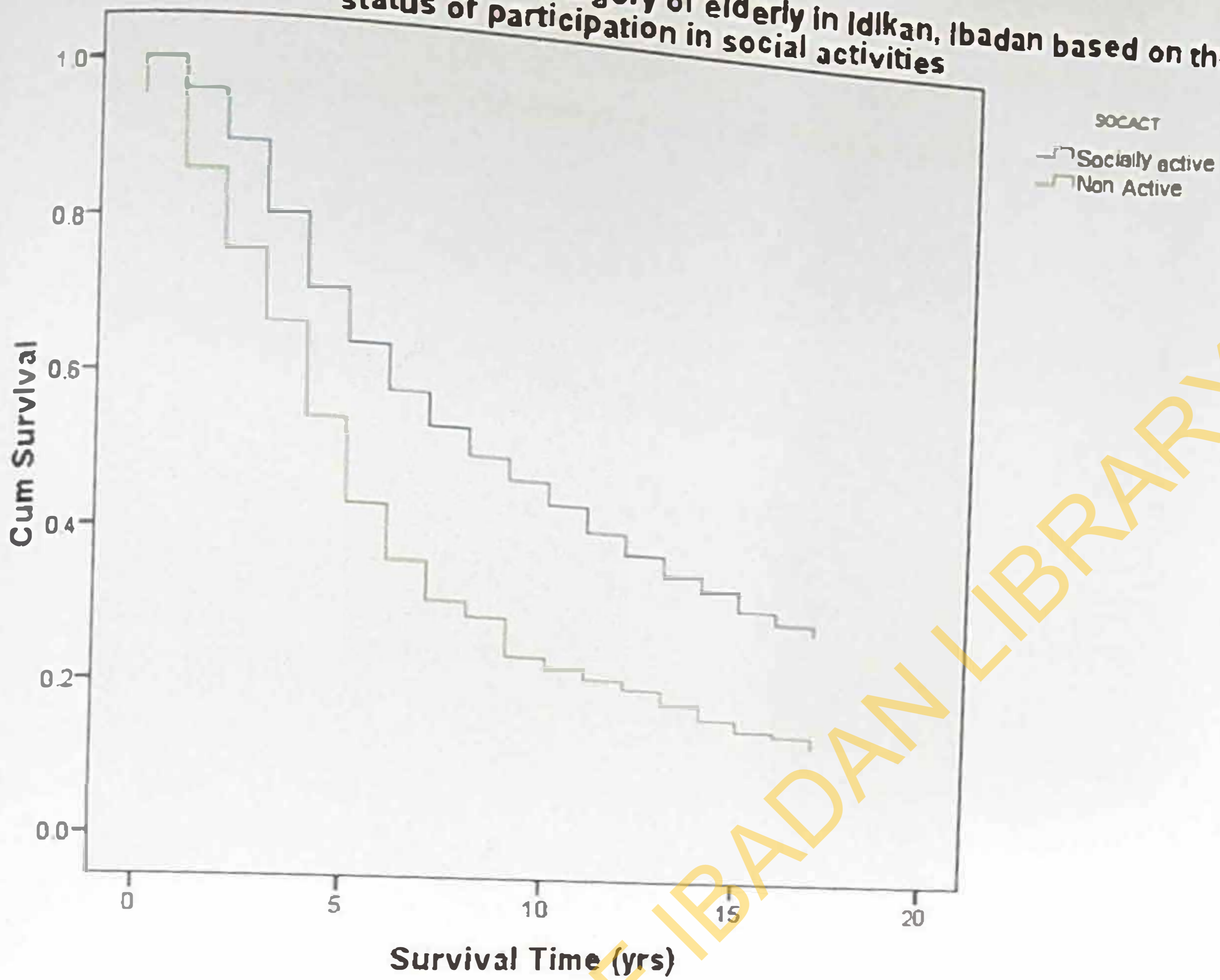

Table 5: Medians Survival Time by status of involvement in social activity and the overall population of the elderly

Categories of Elderly by their status of participation in Social Activity	Median		95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound
Socially active	9.000	.318	8.377	9.623
Non Active	5.000	.223	4.563	5.437
Overall	8.000	.275	7.462	8.538

Table 6: Test of equality of survival distributions for the two categories of elderly in Idikan, Ibadan based on their status of participation in social activities

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	84.369	1	.000

52

Table 7 presented the estimate of coefficient and their odd ratio in the cox regression analysis model of the elderly in Idikan. The result showed that participation in social activities, gender, age, literacy status, presence of chronic disease (s) and cognitive status at the baseline were statistically associated with survival after adjusting for other variables with p-value ≤ 0.05. Living arrangement, alcohol consumption and smoking were not significant.

53

Table 7: Cox regression model analysis result of the elderly in Idikan

Variables in the Equation	B	SE	Sig.	Exp(B)	95.0% CI for Exp(B)	
Status of Participation in Social Activities					Lower	Upper
Socially Active						
Non Active	.431	.064	.000	1.539	1.358	1.743
Gender						
Female						
Male	.351	.061	.000	1.421	1.260	1.602
Age group						
65 to 74 years			.000			
75 to 84 years	.454	.059	.000	1.575	1.402	1.770
85+	.714	.080	.000	2.041	1.746	2.386
Literacy Status						
Can read and write						
Cannot read and write	.164	.071	.021	1.179	1.025	1.356
Presence of chronic disease						
No						
Yes	.179	.080	.026	1.195	1.022	1.398
House Composition						
Lives Alone						
Do not live alone	.357	.221	.107	1.428	.925	2.205
Ever Drinks Alcohol						
No						
Yes	.046	.065	.481	.955	.840	1.086
Ever Smokes						
No						
Yes	.051	.065	.431	1.052	.927	1.195
Cognitive Status			.000			
Good Performance			.001	1.240	1.089	1.413
Intermediate Performance	.215	.066	.000	1.297	1.156	1.455
Poor Performance	.260	.059				

54

CHAPTER 5

5.0. DISCUSSION, CONCLUSION AND RECOMMENDATION

5.1. DISCUSSION

The study showed that there was a female preponderance in the elderly composition. This finding was consistent with, the international sex ratio of more males than females in every five years age bracket until 40-45years of age, after which more females than males. Also the NDHS, 2008 reported a sex ratio of 1:2 for male and female respectively in the elderly population. Another study also reported that male sex was associated with reduced survival (Anthony Perkins et al, 2002).

The poor literacy level of 15.2% observed was not accidental. These subjects were given birth to in the primitive age of early 1920s when the Yorubas practiced polygamy faithfully to give birth to children with a primary objective of using them on the farm to assist in increasing farm produce.

The fact that only 1.2% lived alone, was only a reflection of the Yoruba culture where people lived as extended family. The nature of the Idikian community dwellers who were mainly Muslims by religion (which supports the practice of polygamist as a type of marriage), did not encourage living alone. The fact that a higher proportion of the population was more socially active was also a reflection of the Yoruba ways of life who are well known for parties (funeral, marriage, naming of new born babies, celebrating freedom from the learning of apprentice work), attending meetings (political meetings, church meetings, neighborhood meetings, town meetings, society meetings), playing games in the evening (such as draft, cards, mancala known as 'ayo-opon' in Yoruba language), going to visit friends, relatives or neighbours, and many others. It was even reported in the news that Yorubas are the happiest people in the world.

55

In this investigation, it also appeared that participation in social activity decreases with increase in age. Studies have shown that daily activities of persons aged 70 and older are mainly household chores; more of productive than consumptive activities (Heiner Maier and Petra L. Klumb, 2007).

It also appeared that the females are more socially active than their male counterparts. In reality, the level of engagement of women in society events is obviously higher than that of men such as attending churches, eagerness to wear 'Aso Ebi' and 'Anko' to attend parties, get together for chat, talk and more. Also the females and those with good cognitive performance at the baseline also had better chances of survival with one and a half times likely to live longer as well.

Those not living alone were more socially active than those living alone because they were living among people and would definitely had things doing together. Also, it appeared in this study that those who had poor cognition were less socially active than others. This can be well understood because poor cognition itself is a mental disorder that hinders reasoning, causes depression. Again, engagement in social activities thus affects mental and physical health and, ultimately, survival (Heiner Maier and Petra Klumb, 2007).

The study revealed that gender-being a male sex, increasing age, presence of chronic disease, decreasing cognitive functioning (with high probability of being demented) were associated with increased risk of mortality. This was in consonance with the other findings that being a male sex, increasing age, presences of chronic disease, decreasing cognitive functioning were associated with reduced survival (Anthony Perkinser et al, 2002). In this study, literacy status was also associated with survival while house composition, smoking, alcohol drinking were not associated with increased risk of mortality after adjusting for other variables.

56

It was observed without adjusting for any potential confounder yet that, most of the 'not socially active' had died. Also, participation in social activity appeared effective against death at all points of follow-up and more effective during follow-up than it was early on. The protective effect of social activities against the mortality of the elderly population had been shown in research using variables related to specific aspects of social activities (Hanson et al, 1989; Sabin et al, 1993, Yasuda et al, 1997; Brown, 2003; Giles, 2005]. The study had shown that those who were socially active were almost one and a half times more likely to live longer than those who were not socially active after adjusting for confounding effects. In consistence with this, some studies had shown that participation in social activity prolong life (Evans et al., 1991; Heeren et al., 1992).

It was observed without adjusting for any potential confounder yet that, most of the 'not socially active' had died. Also, participation in social activity appeared effective against death at all points of follow-up and more effective during follow-up than it was early on. The protective effect of social activities against the mortality of the elderly population had been shown in research using variables related to specific aspects of social activities (Hanson et al, 1989; Sabin et al, 1993, Yasuda et al, 1997; Brown, 2003; Giles, 2005]. The study had shown that those who were socially active were almost one and a half times more likely to live longer than those who were not socially active after adjusting for confounding effects. In consistence with this, some studies had shown that participation in social activity prolong life (Evans et al., 1991; Heeren et al., 1992).

57

5.2 CONCLUSION & RECCOMMENDATION

In conclusion, the greatest survival benefit was achieved with being socially active and not just being active at home alone. This work had been able to contribute to the body of knowledge that participation in social activity is a preventive measure against early death. It contributes considerably to achieving longevity.

These highlights the need to consider formulating programs surrounded with social activities programs for elderly people in the society, aim at promoting longer life, including making advocacy to policy makers to build support for this cause. And for the existing programs in the society, it is necessary to promote and preserve them against losing their cultural values.

REFERENCES

Angel Rodriguez-Laso, Maria Victoria Zunzunegui and Angel Otero (2007): The effect of social relationships on survival in elderly residents of a Southern European community: a cohort study. BMC Geriatrics 2007, 7: 19.

Anthony J. Perkins, Siu L. Hui, AdesolaOgunniyi, OyewusiGureje, OlusegunBaiyewu, Frederick W. Unverzagt, SujuanGao, Kathleen S. Hall, Beverly S. Musick, Hugh C. Hendrie (2002): Risk of mortality for dementia in a developing country: the Yoruba in Nigeria. International Journal of Geriatric Psychiatry, Vol.17 Iss.6: pg. 566–573.

AyeniOlusola (2006): Unpublish lecture note on Survival Analysis.

Baltes, Margret M., Ineke Maas, Hans-Ulrich Wilms, Markus Borchelt, and Todd Little (1999): "Everyday competence in old and very old age: Theoretical considerations and empirical findings." Pg. 384-402 in *The Berlin Aging Study: Aging from 70 to 100*, edited by Paul B. Baltes and Karl Ulrich Mayer. New York: Cambridge University Press.

Bamgboye E.A. (2006): Unpublished lecture note on Introduction to Research Methodology.

Barclay L.L. Femcor A., Blass J.P. Sansone V. (1985): Survival in Alzheimer's disease and vascular dementia. Neurology 1985, Vol.35: pg. 834-840.

Batdereschi M.D., Carlo A., Maggi S. et al (1999): Dementia is a major predictor of death among the Italian longitudinal study on Aging. Neurology, Vol 52: pg. 709-713.

Betty R. Kirlwood and Jonathan A.C. Sterne (Second Edition, 2003): Essential Medical Statistics.

Brown S.L., Nesse R.M., Vinokur A.D., Smith D.M. (2003): Providing social support may be more beneficial than receiving it: Results from a prospective study of mortality. Psychology Science, Vol. 14: pg. 320-327.

Burns A.Lewis G., Jacby R., Levy R. (1991): Factors affecting survival in Alzheimer's disease. Psychology Medicine, Vol.21: pg. 363-370.

Cantor, N., and Sanderson, C. A. (1999): Life task participation and well-being: The importance of taking part in daily life. In D. Kahneman& E. Diener (Eds.), Well-being: The foundations of hedonic psychology (pg. 230-243). New York: Russell Sage Foundation.

Clark, L. A., and Watson, D. (1988): Mood and the mundane: Relations between daily life events and self-reported mood. Journal of Personality and Social Psychology, Vol. 54: pg. 296-308.

David G.Kleinbaum: SurvivalAnalysis. A self-learning Text.

Evans D.A., Smith L.A., Scherr P.A., Albert M.S., Funkenstein H.H., Herbert L.E.(1991): Risk of death from Alzheimer's disease in a community population of older persons. American Journal of Epidemiology, Vol. 134: pg. 403-413.

Giles L.C., Glonek G.F.V., Luszcz M.A., Andrews G.R. (2005): Effect of social networks on 10 year survival in very old Australians: the Australian longitudinal study of aging. Journal of Epidemiology and Community Health, Vol. 59: pg. 574-579.

Glass, T. A., de Leon, C. M., Marottoli, R. A., and Berkman, L. F. (1999). Population based study of social and productive activities as predictors of survival among elderly Americans. British Medical Journal, Vol. 319, pg. 478-483.

Hanson B.S., Isacsson S., Janzon L., Lindell S. (1989): Social network and social support influence mortality in elderly men. The prospective population study of "Men Born in 1914", Malmö, Sweden. America Journal of Epidemiology, Vol. 130: pg. 100-111.

Hendricks, C. D., and Hendricks, J. (1998): Aging in mass society: Myths and realities. Reading, MA: Addison Wesley.

Heyman A., Wilkinson W.E., Hurwitz B.J. et al. (1987): Early-Onset Alzheimer's disease: Clinical predictors of institutionalization and death. Neurology, Vol. 37: pg. 980-984

Hodkinson, H.M. (1972): Evaluation of a mental test scores for assessment of mental impairment in the elderly. Age Ageing: Vol. 1: pg. 233-238.

House, J. S., Landis, K. L., and Umberson, D. (1988): Social relationships and health. Science, Vol. 241: pg. 540-545.

Hugh C. Hendrie, Benjamin O. Osuntokun, Kathleen S. Hall, AdesolaO.Ogunniyi, Siu L. Hui, Frederick W.Unverzagt, OyewusiGureje, Cynthia A. Rodenberg, OlusegunBaiyewu, Beverly S. Musick, AbiodunAdeyinka, Martin R. Farlow, S.O Oluwole, Andrew Class, O.Komolafe, Allison Brashear, Vicki Burdine (1995): Prevalence of Alzheimer's Disease and Dementia in Two Communities: Nigerian Africans and African America. America Journal of Psychiatry, Vol. 152: pg. 1485-1492.

Ibadan-Indianapolis Dementia Project Database, 1992-1993.

Jagger C., Clarke M., Stone A. (1995): Predictors of Survivor with Alzheimer's disease's: A community based study. Psychology Medicine, Vol. 25: pg. 171-177

Kaplan, E.L. and Meier, P. (1958): Non parametric estimation from incomplete observations. America Statistical Association Journal, Vol. 53: pg. 457-48.

Kaszniak A.W., Fox J., Gardell D.L., Garron P.C., Huckman M.S., Ranisay R.G. (1978): Predictors of Mortality in Prehensile and Senile dementia. Annual Neurology, Vol. 3: pg. 246-252.

Katzman R., Hill L.R., Yu E.S. et al (1994): The malignancy of dementia. Predictors of Mortality in Clinical diagnosed dementia in a population survey of shanghai, China. Arch. Neuro, Vol. 151: pg. 1220-1225.

Lawton, M. P., Moss M., and Fulcomer M. (1986-87): "Objective and subjective uses of time by older people." International Journal of Aging and Human Development, Vol. 24: pg. 171-188.

Lemon, B. W., Bengtson, V. L., and Peterson, J. A. (1972): An exploration of the activity theory of aging: Activity types and life satisfaction among in-movers to a retirement community. Journal of Gerontology, Vol. 27: pg. 511-523.

House, J. S., Landis, K. L., and Umberson, D. (1988): Social relationships and health. Science, Vol. 241: pg. 540-545.

Hugh C. Hendrie, Benjamin O. Osuntokun, Kathleen S. Hall, AdesolaO.Ogunniyi, Siu L. Hui, Frederick W.Unverzagt, OyewusiGureje, Cynthia A. Rodenberg, OlusegunBaiyewu, Beverly S. Musick, AbiodunAdeyinka, Martin R. Farlow, S.O Oluwole, Andrew Class, O.Komolafe, Allison Brashear, Vicki Burdine (1995): Prevalence of Alzheimer's Disease and Dementia in Two Communities: Nigerian Africans and African America. America Journal of Psychiatry, Vol. 152: pg. 1485-1492.

Ibadan-Indianapolis Dementia Project Database, 1992-1993.

Jagger C., Clarke M., Stone A. (1995): Predictors of Survivor with Alzheimer's disease's: A community based study. Psychology Medicine, Vol. 25: pg. 171-177

Kaplan, E.L. and Meier, P. (1958): Non parametric estimation from incomplete observations. America Statistical Association Journal, Vol. 53: pg. 457-48.

Kaszniak A.W., Fox J., Gardell D.L., Garron P.C., Huckman M.S., Ranisay R.G. (1978): Predictors of Mortality in Prehensile and Senile dementia. Annual Neurology, Vol. 3: pg. 246-252.

Katzman R., Hill L.R., Yu E.S. et al (1994): The malignancy of dementia. Predictors of Mortality in Clinical diagnosed dementia in a population survey of shanghai, China. Arch. Neuro, Vol. 151: pg. 1220-1225.

Lawton, M. P., Moss M., and Fulcomer M. (1986-87): "Objective and subjective uses of time by older people." International Journal of Aging and Human Development, Vol. 24: pg. 171-188.

Lemon, B. W., Bengtson, V. L., and Peterson, J. A. (1972): An exploration of the activity theory of aging: Activity types and life satisfaction among in-movers to a retirement community. Journal of Gerontology, Vol. 27: pg. 511-523.

61

Lennartsson, C., and Silverstein, M. (2001): Does engagement with life enhance survival of elderly people in Sweden? The role of social and leisure activities. Journals of Gerontology,: Social Sciences, 25: S335-S342.

Longino, C.F. Jr., Kart, C.S. (1982): Explicating activity theory: a formal replication. Journal of Gerontology, Vol. 37, Iss.6: pg. 713-22.

McEwen, B. S. (1998): Protective and damaging effects of stress mediators. New England Journal of Medicine, Vol. 338: pg. 171-179.

Musick, M. A., Herzog, A. R., and House, J. S. (1999): Volunteering and mortality among older adults: Findings from a national sample. Journal of Gerontology: Social Sciences, 54, S173-S180.

Nigeria Demographic Health Survey Report, 2008.

OlusholaIbiyemi et al (2011): Assessment of Atraumatic Restorative treatment on the permanent dentition in primary care setting in Nigeria. International Dental Journal: Vol.61, Issue 1: pg. 2-6.

Oxford Advanced Learners Dictionary, 7th edition.

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. and Smith, P.G. (1977): Design and analysis of randomized clinical trials requiring prolonged observation of each patient. British Journal of Cancer, Vol. 35: pg. 1-39.

Petral L. Klumb, Heiner Maier (2007): Daily Activities and Survival at Older Ages. Journal of Aging and Health, Vol. 19, Iss.4: pg. 594-611

Reid, M. (1934): Economics of household production. New York: Wiley.

Reis, H.T., Sheldon, K.M., Gable, S.L., Roscoe, J., and Ryan, R.M. (2000): Daily well-being: The role of autonomy, competence, and relatedness. Personality & Social Psychology Bulletin, Vol. 26, Iss.4: pg. 419-435.

Sabin E.P. (1993): Social relationships and mortality among the elderly. Journal of Gerontology 1993, Vol. 12: pg. 44-60.

Sapolsky, R. M. (1993): Potential behavioral modification of glucocorticoid damage to the hipgocampus. Behavioral Brain Research, Vol. 57: pg. 175-182.

Seeman T.E., Berkman L.F., Kohout F., Lacroix A., Glynn R., Blazer D. (1993): Intercommunity variations in the association between social ties and mortality in the elderly. A comparative analysis of three communities. Annual Epidemiology, Vol. 3: pg. 325-35.

United Nation Human Development Report, 2010.

Van Dijk P.T.M., Dippel D.W.T., Habbema J.D.F. (1991): Survival of patients with dementia. Annual Geriatrics Soc., Vol. 39: 603-610.

Walsh J.S., Welsh G., Larson E.B. (1991): Survival of outpatients with Alzheimer type of dementia. America International Medicine, Vol. 21: pg. 363-370.

Welin, Larsson L., Svardsudd B., Tibblin K., B., Tibblin, G. (1992): Social network and activities in relation to mortality from cardiovascular diseases, cancer and other causes: A 12 year follow up of the study of men born in 1913 and 1923. Journal of Epidemiology and Community Health, Vol. 46, Iss.2: pg. 127-132.

Yasuda N., Zimmerman S.I., Hawkes W., Fredman L., Hebel J.R., Magaziner J. (1997): Relation of social network characteristics to 5-year mortality among young-old versus old-old white women in an urban community. America Journal of Epidemiology, Vol. 145: pg. 516-523.

World Health Organization Report, 1991.