ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

FACULTY OF SCIENCE

DEPARTEMANT OF STATISTICS

DETERMINATION OF SOME FACTORS THAT AFFECT THE SURVIVAL RATE OF SCHIZOPHRENIA PATIENTS: THE CASE OF BUTAJIRA-ETHIOPIA

BY

Dereje Waktola

A Thesis submitted to the office of Graduate programs of Addis Ababa University in partial fulfillment of the requirement for the Degree of Master of Science in Statistics

JULY 2008

ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

FACULTY OF SCIENCE

DEPARTEMANT OF STATISTICS

DETERMINATION OF SOME FACTORS THAT AFFECT THE SURVIVAL RATE OF SCHIZOPHRENIA PATIENTS: THE CASE OF BUTAJIRA-ETHIOPIA

BY

Dereje Waktola

Approved by the Board of Examiners:

Silesh Fanta — **Department Head** — **signature**

Fentaw Abegaz — **Internal Examiner** — **signature**

Emmanuel G. Yohannes — **External Examiner** — **signature**

DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any University and all sources of material used for thesis have been duly acknowledged.

Name: Dereje Waktola

Signiture: _____

Place: Faculty of Science, Addis Ababa University

Date: July, 2008

This thesis has been submitted for examination with my approval as a University advisor.

_____

Professor Eshetu Wencheko

TABLE OF CONTENTS

ACKNOWLEDGEMENT.....i

ABSTRACT.....ii

Content

Pages

CHAPTER ONE.....1

1.1 Background of the Study.....1

1.2 Statement of the problem.....4

1.3 Objectives of the Study.....5

1.4 Significance of the Study.....6

1.5 Limitations of the Study.....6

CHAPTER TWO: Literature Review 7

2.1 The general literature about the model used in the study.....7

2.2 The literature specific to the variable used in the study.....10

2.2.1 Gender, age, and marital status.....10

2.2.2 Substance abuse, physical illness, duration of illness, and inability.....12

CHAPTER THREE: Data and Methodology 15

3.1 Data.....15

3.2 Variables in the Study.....16

3.2.1 The Response Variable 16

3.2.2 Predictor variables.....17

3.3 Methodology.....17

3.3.1 The Proportional hazards model.....25

3.3.2 Estimation of the parameters of proportional hazard model.....28

3.3.3 Assessing the Goodness of fit of the Model.....34

CHAPTER FOUR: Data Analysis... 39

4.1 Descriptive analysis 39

4.2 Univariate analysis.....40

4.3 Estimation of model parameters.....41

4.4 Test for the significance of model parameters.....43

4.5 Tests for proportional hazards 47

4.6 Test for the functional form of the model.....60

CHAPTER FIVE: Conclusions and Recommendations 62

5.1 Conclusions 62

5.2 Recommendations.....63

REFERENCES.....65

APPENDIX.....75

ACKNOWLEDGEMENT

I wish to express my sincere and heartfelt gratitude to my cooperative adviser, professor Eshetu Wencheko for his unlimited constructive advice and comments, substantial amount of constructive criticism for all he is worth and encouragement which contributed to the successful completion of the study.

My appreciation and thanks are extended to other people who have contributed to this thesis directly or indirectly. All the members of Butajira project on the course and outcome of mental disorder, in providing all the necessary information. Of these members I would like to forward my appreciation to Dr. Solomon Tefere, Dr. Tefere Boyera and Ato Girmay Medhin for their good cooperation in medical and psychiatric issue.

Finally, I would like to thank the Department of Statistics for materials provided and allowing me to use the facilities. And lastly but not the least I would like to thank my family, friends for their special and continuous cooperation and encouragement throughout my study period.

Abstract

The mental illness, schizophrenia, is a long-term brain disorder which interferes with a person's ability of thinking, managing emotions, making decisions and relating to other people. Schizophrenic patients have constantly been found to have a higher mortality rate. The aim of this research paper is therefore to make a contribution towards the overview of patients with schizophrenia and thereby to identify important factors that have a significant effect on the survival status of those patients in a five-year follow up. Proportional hazards model was employed to identify the covariates that have a statistically significant effect on the survival longevity of schizophrenic patients. The estimation of the model parameters is done by partial maximum likelihood procedures. Computer programs were prepared to obtain the partial maximum likelihood estimates for the underlying parameters in the model. The study had shown that marital status, substance use and duration of illness have a significant effect on the survival longevity of the patients in Butajira project.

Chapter One

Introduction

1.1 Background of the Study

An estimated 400 million people suffer from mental or neurological disorders or from psychosocial problems such as those related to alcohol and drug abuse. The World Health Report (Brundland, 2001) on mental health states that mental health-neglected for far too long is crucial to the overall well-being of individuals, societies, and countries and must be universally regarded in a new light. This study focuses on one specific mental health problem namely, schizophrenia. Schizophrenia is the most persistent and disabling of the major mental illnesses, affecting one in 100 people worldwide at some time in their life, with an annual new incidence of about 15-30 per 100,000. There are as many as 45-50 million people with schizophrenia worldwide, more than 33 million of them in developing countries (Schizophrenia information).

The word schizophrenia which translates roughly as "splitting of mind" and comes from the Greek roots schizein ("to split") and phren-("mind") was coined by Eugen Bleuler in 1908 and was intended to describe schizophrenia as the separation of functioning between personality, thinking, memory, and perception.

Schizophrenia is a serious brain disorder that disturbs the interaction of mental functions in various ways. Schizophrenia patients experience progressive personality changes and a breakdown in their relationships with the outside world. People who suffer from schizophrenia can have a very broad range of symptoms which cause great distress to themselves and their families. They have disorganized and abnormal thinking, behavior and language and become emotionally unresponsive.

Schizophrenia is a long-term brain disorder which interferes with a person's ability of thinking, managing emotions, making decisions and relating to other people. Schizophrenic

patients may see or hear things that don't exist, speak in strong confusing ways, and believe that others are trying to harm them, or feel like they are being constantly watched.

In some people, schizophrenia appears suddenly and without warning. But for most, it comes on slowly, with subtle warning and gradual decline in functioning long before. They start to isolate themselves, start neglecting their appearance, and show a general difference to life. Among the most common early warning signs of schizophrenia are social withdrawals, suspiciousness, deterioration of personal hygiene, inappropriate laugh or crying, depression, irrational statements, and so on. The symptoms that characterize schizophrenia are delusions (false beliefs that cannot be corrected by reason), hallucinations (usually in the form of non-existent voices), disorganized speech, disorganized behavior, and the negative symptom that refers to the absence of normal behavior found in mentally healthy individuals.

Studies had shown that schizophrenia usually occurs equally in males and females although typically appears earlier in men with the peak age of onset being 20-28 years and 26-32 years for females, as they are beginning to realize their potential and contribute more and more to economic development of their country. Most people with schizophrenia become ill at the age they would be making career choices, undergoing training and forming adult relationships. As the result, they often lack social and work skills.

We do not yet understand what causes schizophrenia. Scientists generally agree that schizophrenia is a group of conditions rather than one simple disease and may therefore be found to have several causes. It is generally accepted by researchers that differences in the brain-chemical or structural, or both-may play a part in the disorder.

Genetic research suggest that while no one gene has been found for schizophrenia, several genes may cause a predisposition that can be triggered by certain life events. A person is more likely to get schizophrenia from his/her parents.

Environmental factors may also play a role. For example, infants who have birth trauma (disasters shock, disturbance) may be at a greater risk of developing schizophrenia. There are other socio-economic factors like poverty that can trigger the onset of schizophrenia. Other factors like stress after loss of a job and substance abuse are assumed to be causes of schizophrenia.

The diagnosis for schizophrenia is based on the self-report experience of the person as well as abnormalities in behavior reported by family members, friend or co-workers, followed by secondary and other supportive signs observed by a psychiatrist, social worker, clinical psychologist or other clinical assessment. The most widely used criteria for diagnosing schizophrenia are from the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, the current version being DSM-TV-TR, and the World Health Organization's International Statistical Classification of Diseases and Related Health Problems-currently the ICD-10. Based on those diagnostic criteria a person can be checked whether he or she is suffering from schizophrenia or not. The World Health Organization has developed the tool Schedules for Clinical Assessment in Neuropsychiatry (SCAN) which can be used for diagnosing a number of psychiatric conditions, including schizophrenia.

Although there is an attitudinal improvement from time to time and with ongoing educational improvement of the society, negative attitudes towards people with such kinds of mental health problems which may often be derived from public misconceptions of what the condition is, and the symptoms associated with it are well documented. These attitudes lead to discriminations in many domains, including at work place and housing, and to rejection by family and friends. They can also lead to anticipated and actual discrimination and internalized stigma to decrease life satisfaction and self confidence, and to increase alcohol use, further social isolation, the inability to work, the loss of jobs or education and subsequently income. As a consequence of these many schizophrenic patients show symptoms of depression, appearing withdrawn from social interaction and they are unable to cope with life. For such patients suicide is seen as the only solution which in turn results in the increased mortality of schizophrenic patients compared with the general population.

When we come to the treatment of schizophrenia, despite extensive research there is still no known "cure" for schizophrenia. However, the outlook for people with schizophrenia has improved greatly in the last few decades and many people can be treated outside hospital and live within the community for most of their lives. Although schizophrenia is a chronic disorder, help is available for the patients with social support, counseling, medication, and so on. Many people with schizophrenia are able to function independently and live satisfying life.

1.2 Statement of the problem

In low-income countries where malnutrition and infections diseases are common, mental disorders or brain conditions are given very low priority. Not only in the developing nations, but also in almost all half of the countries in the world there is no explicit mental health policy and there is no program for coping with the rising tide of brain related disabilities. There are insufficient numbers of skilled policy makers, managers and clinicians and the mental health budgets are inadequate. In most of the world, mental health services are inaccessible or of poor quality. The world health organization (World Health Organization, 1998) reported that, of the countries of the world surveyed: 41% have no mental health policy, 25% have no legislation on mental health, 28% have no separate budget on the mental health, 41% do not have treatment facilities for the severe mental disorders in primary health care, 37% have no community health care facilities, and only about 65% of the beds for mental healthy care are in the mental hospitals.

The professional and public awareness of the real burden of mental disorders and the cost in human, social and economic terms is not yet recognized. Furthermore there are many barriers which prevent millions of sufferers from receiving the treatment they need and deserve. Most people who would benefit from psychiatric treatment and rehabilitation do not have affordable access to such services. Leadership-at all levels-for mental health system development needs to be greatly strengthened.

A 2001 Institute of Medicine report on mental illness in developing countries found that in 1990, over two-thirds of people with schizophrenia in these countries were not receiving any treatment. While many patients with schizophrenia respond to treatments, there are only few developing countries who are taking steps to provide care.

Patients with schizophrenia have constantly been found to have a high rate of natural and unnatural mortality (Ponnudurai et al., 2006). Furthermore an international pilot study of schizophrenia (IPSS; World Health Organization, 1979), the Determinants of outcome of Severe Mental Disorders (DoSMED; Jablensky et al., 1992), International study of schizophrenia (ISOS; Harrison et al., 2001) had shown that the rate of experiencing an event by schizophrenic patients are more likely in low-and middle-income countries. Patel et al., (2006) also studied that high mortality rate have been reported from low- and middle-income countries.

The cause of excess mortality due to schizophrenia is a multifactorial issue and thus somewhat difficult to explain easily. However, determinants of factors that have impact on the longevity of patients with schizophrenia have to be intensively studied.

The aim of this paper is therefore to make a contribution towards the general overview of patients with schizophrenia and thereby to identify important factors that have a significant role on the survival status of those patients using the statistical method called survival analysis. Such analysis usually attempts to answer questions such as: What is the fraction of a population with schizophrenia which will survive past a certain time? What are the important predictor variables that do have a direct or an indirect effect on the survival status of populations with schizophrenia? How do particular circumstances or characteristics increase or decrease the odds of the survival?

1.3 Objectives of the Study

Mortality in patients with schizophrenia is consistently higher than in the general population. Patients with schizophrenia had an increased mortality both from natural causes (cancer, heart

disease, respiratory disease) and unnatural causes (accident, homicide, suicide) (Brown, et.al, 2000). Although a large body of evidence has been accumulated in assessing the excess mortality of patients with schizophrenia, the understanding of the factors that enhance the mortality rate in Ethiopia is not available. The general objective of this research paper is therefore to assess the effect of multiple covariates on survival longevity of schizophrenic patients.

Specific objectives:

- ✓ To demonstrate the technique for analysis of lifetime data.
- ✓ To demonstrate the application of statistics in medical areas especially in psychiatry.

1.4 Significance of the Study

- The results/findings of the study may serve as a preliminary study for the assessment of patients with schizophrenia.
- The results/finding of the study may be used as a basis to create a general awareness of the people towards schizophrenia.
- Based on the results of the study we forward the way that might help to increase the life span of schizophrenia patients.

1.5 Limitations of the Study

The study has different limitations that are assumed not to have a considerable effect on the outcome of the work. The major limitation of this study goes with the problems related to the use of secondary data. As different literatures pointed out, there are different factors that are assumed to have impacts on the mortality of schizophrenic patients. The quality of social support,

numbers of relapses, compliance with treatment, functional status before and after the onset of the illness are some factors that could have impacts. However, we did not get data on these variables to include in the analysis. This may make the study somewhat incomplete.

The second limitation is the controversy about measuring the response variable time. If time is measured by a counter variable in units of time, the Cox model assumes that the hazard rate increases linearly with time, conditional on the covariates in the model. It is possible, however, for the time variable to be a logarithmic or some other function of the counter. But the problem is that the significance or non-significance of the covariates in the model may vary by the type of the time variable used which consequently affects the interpretations and inferences made based on the study results. Smaller time units provide more time intervals, which in turn increase the power of the Cox model (less chance of type II error: thinking there is no relation of the modeled variable to the response when in fact there is). Contrary to this larger time units provide fewer time intervals reducing again the power of the Cox model. Non-existence of appropriate time units may create a problem in the analysis. Furthermore the use of time unit in this case seems arbitrary and not reasonable resulting in different interpretation and conclusions from same data set.

Chapter Two

Literature Review

2.1 The general literature about the model used in the study

The origin of survival analysis goes back to the time when mortality tables were introduced. However, it was not until World War II that a new era of survival analysis emerged. This new era was stimulated by interest in reliability (or failure time) of military equipment. At the end of the war, this newly developed statistical method emerging from strict mortality data research to

failure time research, quickly spread through private industry to control the reliability of products.

As the use of survival analysis grew, parametric models gave way to non parametric and semi parametric approaches for their appeal in dealing with the ever-growing field of clinical trials in medical research. And consequently the application of survival statistical models to psychiatric problem is improving from day to day. Evereitt (1987) points out that increasingly, more advanced techniques such as log- linear models, logistic regression model, and survival analysis (Cox proportional hazards models), among others are being applied in psychiatric research. The author presents an interesting approach to applying statitistical methodology to psychiatric issue. Everitt is not alone in having written on the subject. Others Garside and Roth (1978), Laska, Seigel, and Meisner (1985), and Pocock (1980) discuss some aspects of statistical methodology as applied to psychiatric research.

As it is known, the data in psychiatric research are not compatible with standard statistical models. Survival models have the capability of handling censored data. (Cox, 1972; Cox and Oakes, 1984; Kalbflesch and Prentice, 1980; Miller, 1981) used survival analysis in modeling human lifetimes. Allison (1984); Blossfeld, Homerie, and Mayer (1989); Heckman and Singer (1985); Tuma and Hannan (1984) also use survival analysis in economics and sociology. Several recent articles also described how survival models can be used to explore specific topics including social interaction (Allison and Liker (1982), Gardner and Griffin (1989)), Organizational behavior (Fichman (1988), Morita, Lee, and Mowday (1989)), clinical trials (Greenhouse, Stangle, and Bromberg (1989)) and the life course (Johnson, (1988); Teachman (1982)).

Fergusson, Horwood, and Shannon (1984) used hazard functions to study the time to marital breakdown after the birth of child. Hazard functions had been also used in studies of time to shift in attentions in classroom (Felmlee and Eder (1983); Felmlee, Eder and Tsui (1985)), time to change decision in the face of irrelevant information from a low- status partner (Hembroff and Myers (1984)), in study of relapse of mental illness (Lavori et al., (1984)), marital dissolutions

(Morgan, Lye, and Condran (1988); Tuma, Hannan, and Groeneveld (1977)) and Human lifetimes (Gross and Clark (1975)).

Proportional hazards modeling is the most frequently used type of the survival analysis modeling in many research areas, having been applied to topics such as smoking relapse (Stevens and Hollis, (1989)), affective disorders (Shapiro et al., (1989)) , childhood family breakdown (Fergusson, Horwood, and Dimond (1985); Fergusson, Horwood, and Shannon (1984)), interruptions in conversation (Dress (1986)), and employee turnover (Morita, et al., (1989), and in medical areas for identification of important covariates that have a significant impact on the response of the interested variables.

Yamoguchi and Kandel (1985, 1987) propose the modeling of proportional hazard models under the condition that the predictor variables can be simultaneously included to the model be it is time-varying and/or time-invariant in studying the risk of premarital pregnancy, the risk of job turnover. Furthermore other researchers pointed out modeling proportional hazard models in situations where an individual at risk is of one of several states provided that the multiple states are mutually exclusive and exhaustive. For example, Zatz (1985) in a study of the final case dispositions of juveniles arrested in Phoenix, Arizona, in which after arrest, each juvenile's case can be disposed of in one, and only one, of five way:(a) dismissal, (b) informal processing, (c) probation, (d) commitment to the department of corrections, (e) remade to adult court as a possible multiple states which are mutually exclusive and exhaustive to be used in the proportional hazards model.

Pulver, et al. (1990) use Cox proportional hazards models to examine the relationship between age at onset of psychosis for possible schizophrenic examination and familiar risk. Suviseari, et al. (1998) use linear mixed models and the Cox proportional hazards model to assess the relationship between burden to the family, age at onset, and outcome in schizophrenic patients.

The literature introduced thus far pointed out the use of statistics for analysis of psychiatric issue, in particular the use of proportional hazards models for the identification of different factors that are assumed to have influence on the survival longevity of schizophrenic patients.

2.2 Specific literature related to the variables used in the study

2.2.1 Gender, age, and marital status

A number of a research had been made to see the impact of different factors on schizophrenic patients. In the study of the major sociodemographic correlates of schizophrenia and their interactions in a rural population of Ethiopia, Derege Kebede et al., (2004) had shown that being male, under 35 years of age, unmarried are factors all associated with schizophrenia independently of each other. The study had also shown that the results obtained are similar to those described for the developed world. Other evidences also demonstrated that older age, not married, and being female are sociodemographic factors that increase the risk of rehospitalization of people with mental illness.

The relationship between marital status and schizophrenia has been repeatedly shown (Mazberg, 1936) that the first psychiatric admission rate for single schizophrenia was 5.4 times greater than for married schizophrenic. Others had also confirmed this association (Munk-Jurgensen, (1987); Nurris, (1956); Odegaard, (1946); Pokurny and Overall, (1970)).

Gove, et al. (1983) show that being married is very strongly associated with experiencing a happy home life, while the three unmarried statuses namely widowed, divorced, and single are associated with a relatively unhappy home life. The study also indicates that home life satisfaction is much stronger for men than for women. Marital status emerges as the powerful predictor of mental health, even though it varies among the genders. Nuttall and Solomon (1970) examined different demographic variables for 259 schizophrenic patients in three Boston area hospitals. They conclude that marital status was one of the most sociodemographic variables for predicting the length of their hospital stay.

Other studies also demonstrate that marriage reduces adverse schizophrenic outcome such as length of hospitalization, social vocational disfunctioning, and the proportion of time at risk during hospitalization (Chapman, Day, and Burnstein, (1961); Marks, (1963); Mason, et al., (1960)). Marital status predicts schizophrenic's length of hospitalization or recovery as well as

state-of-the art prognostic scales (Chapman, et al. (1961); Farina, Germezy, and Barry, (1963); Garfield and Sundland, (1966); Turner, et al. (1970)).

A national health survey of 122,859 (American National Center for Health Statistics,1988) found that married people spend fewer days in bed due to acute illness than singles, that widows have twice as many acute problems as married women, and that divorced women have double the rate of injuries as married people. A number of literature reviewed that marriage consistently shows lower mortality rates than being single, widowed, or divorced.

Some investigators suggest that marriage may be more of stress than support (Foorman and Lloyd, 1986). Some interpersonal relationships may create high levels of psychiatry distress because, rather than providing emotional support, they impose obligation or demand. Robertson (1974) argues that the married state for women often acts as additional strain leading to the development of mental illness. Walker, et al, (1985) study based on 882 psychiatric patients found that being married had more negative consequence for females than males. Other evidence shows that women have more mental illness than man regardless of marital status.

Suvisaari, et al. (1998) assessed the relationship between burden to family, age at onset and outcome of schizophrenia and concluded that high familial loading for schizophrenia is associated with early onset and poor outcome of schizophrenia. In a family history of 366 schizophrenic patients and their 1851 first degree relatives, Pulver, et al. (1990), found a relationship between age at onset of psychosis in the male patients and the risk of schizophrenia in their relative. The study pointed out that the relatives of male schizophrenic patients whose onset of psychosis occurred when they were younger than 17 years of age had an increased risk of schizophrenia when compared with the relatives of male patients with an age at onset greater than 17. Although the above study points out the effect of age at onset of psychosis on familial relationship and the relatives of the psychosis patients, they did not consider the age at onset on the survival rate of the patients himself. The study by Conwell and Caine (2007) show that the mortality rate was significantly higher among people with latter onset of schizophrenia (>45 years) than among those with onset before 45 years of age.

The literature given above indicates that the effect of sociodemographic factors on mental illness (schizophrenia in our case) is still controversial. Furthermore, a number of publications reveal more or less a consistent pattern that married individuals, especially married men, experience less stress than their unmarried counterparts. In fact, studies of mortality and morbidity, on schizophrenic and other psychiatric problems generally support this. However, their impacts on the longevity of the schizophrenic patients are not forwarded yet.

2.2.2 Substance abuse, physical illness, duration of illness, and inability to work

The vast majority of people with schizophrenia misuse substances. Substance abuse is common, often undetected, and adversely affects outcomes in patients with schizophrenia. In the Epidemiologic Catchment Area Study sample of patients in community and institutional setting in the US, 47% of the schizophrenia subjects were found to be of substance dependent. Alcohol, followed by Cocaine and Cannabis, were the drugs most frequently reported by schizophrenia patients.

People with schizophrenia are subjected to a long term effect of antipsychotic medication and have high rate of substance misuse. A number of hypotheses exist on the relationship between substance misuse and established schizophrenia, and reasons why such psychotic may use substances. One hypothesis is that both the choice and use of substances is a method of "self-medication', to treat adverse states induced by either schizophrenia or its treatment (Hall et al., 1979; Dixon et al., 1990). A number of literatures also argue that patients with schizophrenia use substances to reduce psychotic symptoms and alleviate the sedating side effects of neuroleptics.

Psychostimulant drugs may reduce negative symptoms such as apathy and withdrawal, while alcohol and benthodiazepines can temporarily suppress psychotic symptoms; a possibility is that psychostimulant drugs counteract distressing neuroleptic induced extraptyamidal side effects, particularly akathisia (Knudsen and Wilmer, 1984; Miller, et al. 1989). Other most common reasons for the use of substances especially alcohol and other drugs are to reduce tension,

depression, and so on. Bimerew, Sonn, and Kortenbout (2007) studied that schizophrenic patients use alcohol and khat in order to tolerate the severe side effects of the anti-psychotic drugs, to suppress hunger due to shortage of food and to avoid drowsiness.

Excessive substance use by people with schizophrenia has almost the same adverse social, health, economic, and psychiatric consequences as it does for other individuals. However, it has additional serious consequences for multiply handicapped patients. Substances of abuse tend to increase dopaminergic activity, thereby increasing the risk of symptoms exacerbation and relapse and compromising the efficiency of neoroleptic medications. Substance use is also thought to decrease compliance with treatment and often serve as a source of conflict in families, a pernicious circumstance for patients with schizophrenia, who are highly vulnerable to heightened stress (Alan and Carlo 1999).

Drug and alcohol abuse among psychiatric patients is an increasingly recognized problem (Galonter and Castaneda, (1988); Drake and Wallach, (1989); El-Guebaly, (1990); Regier, et al. (1990)). Other studies had shown that between 60% and 90% of people with schizophrenia smoke cigarettes, which has a substantial impact on mortality of this population. Different studies had shown that substance abuse has been associated with refractory psychotic symptoms, depression, poor treatment compliance, relapse, suicide, and human immunodeficiency virus (HIV). Co-occurring substance use disorders, mostly involving alcohol, cannabis or cocaine, occur commonly in patients with schizophrenia and are associated with increasing mortality.

The Epidemiologic Catchment Area study (Regier, et al. (1990)) suggested that lifetime rates of substance abuse among the mentally disordered may be as high as 87% for individuals with antisocial personality disorders, 56% for bipolar mood disorders, 47% for schizophrenics and 32% for depressive disorders in the USA.

Psychiatric patients have high rates of physical illness. Philips (1953) reported that there is an association between mental illness and poor physical health. However, the impacts of the poor physical illness on the survival rate of schizophrenic patients are not yet investigated.

The study based on ten-year follow up of schizophrenic patients (Limosin, et al. 2003) pointed out that gender, drug abuse, suicide attempts, and duration of illness are among the important variable that have significant impact on the suicide risk in schizophrenic patients. Other studies such as the study based on 9156 first admitted schizophrenic patients (Mortensen and Juel, 1993) argue that age, gender, and duration of illness have a significant impact on the mortality of schizophrenic patients.

Unemployment among people in inner London with long-term mental health problems increased from 80 % in 1990 to 92 % in 1999, and the unemployment rates among those with a diagnosis of schizophrenia increased from 88 % in 1990 to 96 % in 1999 (Perkins and Rinaldi, 2002). For many people who have experienced mental health problems, the main barrier to employment is unwillingness on the part of employers to consider them because of their psychiatric history (Manning and White, 1995; Rinaldi and Hill, 2000). Although some patients may have more disabling ongoing problems that can impede work performance, however research indicate that, with ongoing support, a mean of 58 % of people with serious mental problems can achieve competitive employment (Bund et al, 1997, Crowther, et al. 2001).

Bennett (1975) has characterized work as the performance of a task within prescribed limits to achieve goals set by others, who judge and reward and thereby link the individual to society. Without work these links are all too easily lost. The importance of work in enhancing the mental health and quality of life of people who experience mental health problems had also been widely documented by other researchers (Warner, 1994; Rowland and Perkins, 1988; Nehring, et al. 1993; Shepherd, 1984). The ability to work provides social contacts and social support, and as a means of structuring and occupying time and a sense of personal achievement. Thus one can conclude that inability to work have a role on the survival of schizophrenia patients.

The above literature indicated the relationship between mental illness and scociodemographic factors such as marital status, gender, age and other factors such as substance abuse, physical illness, duration of illness, and inability to work. However, the analysis of the impact of those factors on the survival longevity of schizophrenic patients is almost poorly addressed.

Chapter Three

Data and Methodology

3.1 Data

The data were obtained from the study project on the schizophrenia and bipolar disorders course and outcome in the Butajira district, which is located 135 kms south of Addis Ababa. The district is predominately rural with a population of about 227,135 of which over 100,000 (45%) belong to the age group between 15 and 49 years. This age group is the target population. In this district the climate varies between temperatures in the highland and tropical lowlands. The major ethnic group is Gourage. The population is predominantly Muslim. Maize and Enset (false banana) are the main crops in this district. Health facilities are very much limited to towns and close-by sub districts and there was no psychiatric service prior to the establishment of a psychiatric clinic by the schizophrenia and bipolar disorders course and outcome study project.

The study was designed for mental health research by the Department of Community Health in the Faculty of Medicine, Addis Ababa University. For data collection a standardized psychiatric research instrument, the CIDI (Composite International Diagnostic Interview) was translated into Amharic and tested for reliability, feasibility of use and acceptability by respondents in Addis Ababa in 1993. For the data collection a survey that involves the entire 15-49 years old population of the area was conducted using the CIDI. Subjects who were CIDI positive and others who were identified by key informants undergo evaluation by clinicians using another standardized diagnostic instrument, the SCAN (Schedules for Clinical Assessment in Neuropsychiatry) which had also been translated to Amharic. Many other instruments were also used to collect information on key informants, patients' mental and physical disabilities and other aspect of patient's treatment outcomes. The patients identified are being followed up to study the course and outcome of schizophrenia, bipolar disorder and major depression.

The total adult population of the Butajira district, reaching 68,491 individuals with CIDI and key informants had been screened. Over 800 cases of schizophrenia, bipolar disorders and major depression were identified and enrolled for follow up during the five years period 1997 to 2001 to study the outcome about the patients. From those identified patients up on using test with different criteria only about 327 patients where found to be schizophrenic. The survival durations for each of the patients in the study in the follow up period is measured during the number of months he/she is in the study.

The sample data collected on those patients demonstrate some important properties that require use of survival analysis techniques. First, and most importantly, the data are censored, which means the event of interest - death was not observed on all patients. Some of the patients are still alive at the end of the limited observation period. It was known that they survived to the end of the observation period. Furthermore, some patients are lost to follow-up, and are also said to be censored.

In fact, due to such remarkable properties of the data, we have to set some assumptions so that all the sample data can contribute information to the output interest. One important assumption made in the analysis of survival data is that censoring is non informative. That is, censoring occurs randomly, and the patients with censored observations are not different from those with uncensored observations with respect to their chance of experiencing the event of interest in given time period. This was important in allowing each experimental unit to contribute all of the information possible to the model for the amount of time the researcher was able to observe the unit.

3.2 Variables in the Study

3.2.1 The Response Variable

In medical studies the hazard might be death. In industrial studies the hazard might be engine breakdown. A hazard rate at a given time is the probability of a given event occurring in that time period, given survival through all prior time intervals. The time, we consider as response

variable in the survival model literature, measures duration to event defined by status variable (the status variable also called the event or censoring variable is the dependent variable in the Cox regression). This is to say that the response variable in the Cox regression model is a censored survival time represented by variable time and event/death. The survival times are data that measure follow-up of time from a defined starting point to the occurrence of a given event. This observation time has two components that must be carefully defined in the beginning of any survival analysis: the beginning point of the study time and the observation of time to end. In survival analysis, it is the duration of time until the end-point occurs that is the outcome of interest. Normally the time variable is a simple counter unit of time since the start of the series.

3.2.2 Predictor variables

Covariates are the predictor variables in the Cox model. These covariates can be categorical or continuous. They may be also either time dependent or time fixed, a difference of the different classes of the covariates affects how the covariates would be modeled in Cox procedure. The predictor (covariate) variables to be included in the model are:

- Marital status
- Presence of co morbid substance abuse
- Age at onset of illness
- Current age of schizophrenic patients
- Gender
- Duration of illness
- Physical illness
- Years of education

3.3 Methodology

Survival analysis is defined as a branch of statistics which deals with death in biological organisms and failure in mechanical systems. This topic is called reliability theory or reliability analysis in engineering, and duration analysis or duration modeling in economics or sociology.

More generally, survival analysis involves the modeling of time-to-event data. In this context, death or failure is considered as an event from which the name survival analysis and much of its terminology derives, but the ambit or the area of applications of survival analysis is much broader in the survival analysis literature.

Survival analysis examines and models the time it takes for events to occur. Survival analysis is being used with increasing frequency in the medical literature. The term survival analysis applies to techniques in which the data being analyzed represent the time it takes for certain event to occur. For example, the survival time of cancer patients from treatment until death, length of time from operation until failure of the transplanted organ, life time of products until failure (reliability) and so on. In fact, the start and end points for measuring time are chosen according to the context. It may not represent the entire life of the study problem.

The use of survival analysis, as opposed to the use of different statistical method, is most important when some subjects are lost to follow up or when the period of observation is finite such that not all patients experience the event of interest over the study period. Usually we have some individuals for whom there is no time-to-event record. Instead all we know is that during certain period of observation there was no event. In reality such situations can occur due to the following reason:

- A medical study is terminated before some individual experienced the event. This termination could be either due to funding problem or the need to write report.
- Some individuals left the study before they experienced the event. That is, subjects may be lost to follow up either due to moves or refusal to participate.
- Termination of the study could be due to negative side effects of treatments.
- If a study is concerned with one particular cause of death, individuals might die of other causes (death due to unrelated cause) and soon.

Therefore the analysis of survival data requires special techniques since the data are almost always incomplete. The statistical terminology for such data is censoring. Hence censoring is a form of missing data problem which is common in survival analysis and it is considered as a distinct feature of survival data. Survival analysis is well suitable for such data which are very

common in medical research because studies in medical areas have a special feature that follow-up studies could start at observation time and could end before all experimental units had experienced an event. This is to mean that once we started a follow-up with some observation, the analysis on those data may be made before all the experimental unity experience an event-death in our case.

Although investigators follow subjects until they reach a pre-specified endpoint (for example, death of patients, failure of components), this may not be the case in most real situations. Consideration of the censored data is extremely important because even in the best developed studies there will be subjects who choose to quit participating, who move far away to follow, or who will die from some unrelated event. The non-censored survival times are referred to as event times. Methods for survival analysis must account for both censored and non-censored data.

Ideally it is assumed that both the birth and death dates of a subject are known, in which case the lifetime is known. If it is known only that the date of failure/event is after some time/period, this is called right censoring. That is, a right censored case is one for which the censoring event had not yet occurred by the end of the measurement period. If the subject's life time is known to be less than a certain duration, the life time is said to be left censored. That is, a case is left-censored when the time of the failure/event is known only to have occurred before time t. It may also happen that subjects with a life time less than some time may not be observed at all. This is called truncation. Truncation is different from left censoring, since for left censored datum, we know the subject exist; but for truncated datum, we may be completely unaware of the subject.

The prospect of censoring complicates the research design, the distribution theory for the estimators complicated statistical analysis. Thus censoring creates some unusual problem in the analysis of data because such data cannot be handled properly by standard statistical methods. Researchers use different techniques to respond to the complication due to censoring. Among these: Some researchers try design solutions by restricting data collection to uncensored observations (Taber and Proch, 1987). Others try to get analytic solutions by categorizing the outcome and placing the censored observation in one group (Condiatte and Lichtenstein 1981),

deleting the censored observations (Litman, Eiser and Taylor, 1979), and adjusting the censored outcome as a categorical predictor of another outcome that varies over time (Coetho,1984). No one of these techniques is entirely satisfactory.

For years, different researchers recognized the limitation associated with solutions to accommodate censored data. Until recently, however, few alternatives were available. Now new developments in statistical theory accompanied by new development in statistical computing have changed how researchers can study such data. This new method known as survival analysis /event history analysis was developed by biostatisticians modeling human life times (Cox, 1972, Cox and Oakes 1984; Kalbfleish and Prentice, 1980; Miller, 1981) in the medical and biological sciences. And that has been extended by economists and sociologists studying social transitions (Allison, 1995; Blossfeld, Hamerle and Mayer, 1989) as well as in engineering (reliability and failure time analysis).

Once this method was developed in modeling human life time where the target event is death, it has been serving as a powerful methodology that appropriately uses data from all observations and become popularly used in sociology, demography, psychology, economics, political science, and marketing. It does not matter whether the data are uncensored or censored. Data collection can be prospective or retrospective, experimental or observational. Time can be measured continuously or discretely.

There are obviously many potential life models. In some situations there may be reasons to select a particular family of models: the model may fit data on hand well, past experience may have shown the model to give a good description of lifetime distribution from similar populations, there may be a knowledge of the underlying aging or failure process that suggests the validity of the model, and so on. In situations in which no family of models is singled out as being particular appropriate, the choice of the model is frequently made on the basis of considerations such as: the convenience of mathematically handling the model, the statistical methods available in connection with the model and the degree of complication of calculations involved in using the model. Three additional points should be mentioned in connection with the choice of the model. Firstly, for any chosen particular model it has to fit the available data up on

appropriate tests. Second, one should be aware of the consequences of departures from the assumed model on inferences made. Finally, although there is a lot of work based on lifetime distributions of parametric models, there are many situations in which it is desirable to avoid strong assumptions about the model. Nonparametric or distribution-free procedures are important in this case (Lawless, 1982).

In most real situations, survival models are constructed by choosing a basic survival distribution. The choice of survival distribution expresses some particular information about the relation of time and any exogenous variables to survival. It is natural to choose a statistical distribution which has a non-negative support since survival time is non-negative, regardless of its complexity.

The major mathematical complication with survival analysis is that we usually do not have the luxury of waiting until the very last subject has died of or experiencing an event. We normally have the data while some subjects are still alive. Also, some subjects may have moved away, and may be lost to follow up. In both cases, the subjects are known to have survived for some amount of time (up until the time you last saw them) but we do not know how much longer they might ultimately have survived. Several methods have been developed for the analysis of survival data. Some of these are:

a) Descriptive statistics include life tables, survival distribution, and Kaplan-Meier survival function estimation which are used for the estimation of the distribution of survival time from a sample.
b) Univariate statistics involve techniques that are available for comparing the survival in two or more groups. The most common and widely used of this technique is the log-rank test. The Wilcoxon technique is less commonly used because it gives greater weights on events close to zero.
c) The Multivariate Method mostly uses Cox-proportional hazards model. It is considered as the most interesting survival modeling in the interest of examining the relationship between survival and one or more predictors, usually termed covariates in the survival analysis literature. Covariates may be categorical (for example the marital status of an individual in the study etc...) or continuous (for example the

patient's age, weight etc...). In addition the model has the capability of including both time-dependent and time-independent variables.

The Cox-proportional hazards model has the form of regression model. In fact there are many types of regression models. Two of the most popular types of models are the accelerated failure time model (Kalbfleish and Prentice, 1980) and the Cox proportional hazards model (Cox, 1972). Each has its own assumptions on the underling distribution of the survival times. In each case the linear combination of the covariate is assumed to act additively on one of:

i. The log of the survival time-accelerated life model: The model assumes that the shape of the survival curve is the same for all patients, but they move along it faster or slower according to the value of their covariates. On the other hand the accelerated failure time model assumes a parametric form of the effects of the predictor variables and usually assumes a parametric form for the underlying survivor function. This model is widely used in engineering applications, but less frequently used in medical studies. It may also relate to the complications involved in relating the model described above and the usual survival distribution.

ii. The log of the hazard time-proportional hazards model: In this model the effect of the covariates is to increase or decrease the hazard functions by a constant proportion relative to the baseline hazard function. Here also the proportional hazards model assumes a parametric form for the effect of predictor variables, but it allows an unspecified form for the underlying survivor function.

The proportional hazard model is the most popular regression method for analysis of censored survival data. The popularity is because of its:

- Flexible choice of covariates (we can accommodate time varying covariates, time independent covariates, continuous covariate, and discrete covariates).
- Fairly easy to fit.
- Standard software exist (different well known statistical software such as SPSS, SAS, STAT, S-PLUS etc are programmed in such a way that they have the capability to handle proportional hazards model).

- Does not make assumption about the underling survival distribution (does not require the knowledge of the shape of the survival distribution).
- Does not require estimation of the baseline hazards rate, λ_0 (t) to estimate the regression parameters, β's (the primary interest of the model is estimation of the regression parameters β's with no attention on the base line hazards function λ_0(t)).

Unlike parametric models, the proportional hazards model is not based on any assumptions concerning the nature or shape of the underlying survival distribution. This means, we do not have to choose the density function of a parametric distribution of the survival time, making the methods considerably more robust than the parametric models if the other assumptions of the Cox regression are met. Thus in this sense, this model may be a nonparametric method. Instead it requires the assumption that the hazard ratio, also called the hazard function which describes the effect of predictor on survival, is constant over the entire follow up period (Cox 1972). On the basis of this, one can think that the Cox's proportional hazards model is semi parametric; it makes the parametric assumption concerning the effect of the predictors on the hazard function, but makes no assumptions regarding the nature of the hazard function itself.

Due to its dual nature, proportional hazards model, which we also call the Cox-regression model or Cox-proportional hazards model, may be preferred over parametric models. Only when one has strong theoretical reasons for assuming a particular distribution (shape) for the baseline hazard function which is very rare, then a parametric model is preferred. In most cases there is no such strong, clear reason and the more stringent data assumptions of parametric models are not justified, making Cox models the better choice to generate the same information as parametric models without having to make strong data assumptions.

More importantly in many situations, either the form of the true hazard function is unknown or it is complex and more interest centers on the effects of the predictors than the exact nature of the hazard function. In such a situation, Cox's method has the distinct advantage of allowing the shape of the hazard function to be ignored (Geoff and Everitt, 2006).

Through a logarithmic transformation, the proportional hazards model can be expressed in the form of a multiple regression model. From this similarity of the two models some people may bother why the usual multiple regression model cannot be used. That is people may say why we cannot use a multiple linear regression model with time as a response variable for analysis of our survival data.

If we have ever used regression analysis to analyze such data, we face two intractable problems:

- The response variable of interest (survival or failure time) is most likely not normally distributed. It is positively skewed indicating that a few people survive a very long time compared with the majority. Thus, assuming a normal distribution will not be reasonable. This is a serious violation of an assumption of ordinary least square multiple regression. Survival time usually follow an exponential or Weibull distribution. Thus the analysis of time-to-event data requires some other non-standard statistical methods.

- There is a problem of censoring with survival analysis which cannot be handled by usual standard statistical methods such as multiple regression models.

In addition to the fact that a multiple regression model is designed for continuous variables, makeshift solutions to these problems can lead to severe biases. There are also regression methods for use with discrete time variables: If, for example, we dichotomize time into event before or after a fixed time, we could use the logistic regression models, apart from problems caused by censoring. If there were no censoring and if our interest is only in predicting the probability of survival to this fixed point, logistic regression would be an attractive model, since it allows us to concentrate the accuracy of the model on the precise quantity we wish to estimate rather than fitting the complete survival distribution. However, it is not valid to ignore the censored individuals especially when censoring is considered to be non-informative.

Therefore, survival methods are explicitly designed to deal with censoring and time-dependent covariates in a statistically correct way. If the data were not censored, i.e., if all patients died during the observation period and none were lost to follow-up, familiar non parametric statistical techniques such as the Wilcoxon rank sum test and the Mann-Whitney U test could be used to

compare survival times. The ability to properly analyze censored data is the key feature that differentiates survival analysis from other non parametric statistical methods. If common nonparametric techniques were applied to data that contain censored observations, the information that is available from the censored observations would be lost.

However, with the use of the Cox proportional hazard model, an investigator is no longer forced to withdraw the experimental unit and all associated data from the study. Instead, it enables an investigator to analyze incomplete data due to delayed entry or allowing each experimental unit to contribute all of the information possible to the model for the amount of the time the investigator is able to observe the unit.

3.3.1 The Proportional hazards model

The Cox proportional hazards model is a multiple regression method used to evaluate the effect of multiple predictor variables on the survival. This model was proposed by Cox (1972), and has also become to be known as the Cox regression model. The proportional hazards method computes a coefficient for each predictor variable that indicates the direction and degree of flexing that the predictor has on survival. In that sense a value zero means that a variable has no effect on the survival; a positive value indicates that a large value of the variable is associated with greater hazard rate (poorer survival prospect). The model assumes that changing, the predictor variable vector results in a new hazard function that is proportional to the baseline hazard function, and the proportionality constant is a function of the predictor variable vector, independent of time.

We often model the proportional hazards model through the hazard function which measures the risk or proneness to death at time t. The hazard function is defined as the probability that an individual dies or fails to survive or experience an event, if he or she has already survived up to that time point. The hazard or risk of dying may decrease over time, death after an operation, or

increase over time as occurs with normal aging (retirement age), or remain steady (death due to some accidents), or may have a more complex shape. The hazard function for death of human beings has the 'bathtub' shape. Thus, the hazards function can be decreasing, increasing or discontinuous, but the only restriction here is that it should be nonnegative.

Cox (1972) proposed a semi parametric model for the hazard function that allows the addition of predictor variables, or covariates, but keeps the baseline hazards an arbitrary, unspecified, nonnegative function of time. The Cox proportional hazards function for fixed-time covariate Z, which is flexible enough to handle time-dependant covariate, is defined as follows.

The equation of proportional hazards model (Cox, 1972) is

$$\lambda(t;Z) = \lambda_0(t)\exp(\beta' Z) \qquad (1)$$

where $\lambda_0(t)$ is an arbitrary unspecified base-line hazard function for continuous T, which reflects the underlying hazards for subjects with all covariates equal to zero; $\lambda(t;Z)$ represents the hazard function at time t for an individual with covariates Z; $\beta = (\beta_1, \ldots, \beta_p)'$ is a column vector of p regression parameters, which measures the influence of the covariate on the survival, with β_i representing the increase in the log hazards as z_i increase one unit relative to the baseline hazard function $\lambda_0(t)$. That is to say, $\exp(\beta_i)$ gives the relative risk associated with an increase of one unit in z_i, all other predictor variables remaining constant. The vector β is assumed to be the same for all individuals; $Z_i = (Z_{i1}, \ldots, Z_{ip})'$ is a column vector of p measured covariates which is believed to affect the life time of schizophrenic patients and t is the associated failure time.

The survival times of each member of the population is assumed to follow its own hazard function. In such a case, the above model can equivalently and explicitly be written as

$$\lambda_i(t;z_i) = \lambda_0(t)\exp(\beta_1 z_{i1} + \ldots + \beta_p z_{ip}) \qquad (2)$$

$i = 1, \ldots, n$, where n is total numbers of individuals in the study.

While no assumptions are made about the shape of the underlying hazard function, that is no particular form of probability distribution is assumed for the survival times, the model in equation (1) shown above do implies two basic assumptions.

- First, they specify a multiplicative relationship between the underlying hazard function and the log linear function of the covariates. That is to say, hazard functions for all patients are proportional to each other - they have the same basic shape as the base line hazard function. This assumption is also called the proportionality assumption. In practical terms, it is assumed that, given two observations with different values for the predictor variables, the ratio of the hazard functions for those two observations does not depend on time.
- The second assumption is that there is a log-linear relationship between the predictor variables and the underlying hazard function. Less explicitly, it is also assumed that individuals are independent and homogeneous given their covariates (Andersen, 1991). Thus, the Cox model in equation (1) is only valid for the data consistent with the assumption of proportional hazards.

The important features of the above model are: the baseline hazard depends on t, but not on the covariates and secondly the hazard ratio, that is, exp ($\beta' Z$), depends on the covariates, but not on time t. The second feature is in fact what leads us to call equation (1) a proportional hazards model equation and which leads to one of the important assumptions of the proportional hazards model, is the proportionality assumption.

This important assumption of the proportional hazards model can be expressed mathematically as follows. For any two observations i and i' that differ in their z-values, with the corresponding linear predictors called the risk scores or the prognostic index

$$\eta_i = \beta_1 z_{i1} + ... + \beta_p z_{ip} \qquad (3)$$

$$\eta_{i'} = \beta_1 z_{i'1} + \ldots + \beta_p z_{i'p}$$

The hazards ratio for these two observations

$$\frac{\lambda_i(t;z_i)}{\lambda_{i'}(t;z_{i'})} = \frac{\lambda_o(t)\exp(\eta_i)}{\lambda_o(t)\exp(\eta_{i'})} = \frac{\exp(\eta_i)}{\exp(\eta_{i'})} = \exp(\eta_i - \eta_{i'}) \qquad (4)$$

is independent of time t.

This shows that the ratio of the hazard functions for two individuals with different covariates does not vary with time. That is, their ratio is assumed constant over the survival time, thereby not allowing a temporal bias to become an influential player on the endpoint.

3.3.2 Estimation of the parameters of proportional hazard model

One of the biggest advantages of the framework of the Cox proportional hazards model is that we can estimate the vector of parameters β which reflects the effect of covariates without having to make any assumptions about the form of λ_0 (t). In other words, we don't have to assume that λ_0 (t) follows an exponential model, or a Weibull model, or any other particular parametric model. The proportional hazards model is non-parametric in the sense that it involves an unspecified function in the form of an arbitrary base-line hazard function. In consequence, this model is more flexible and an estimate of the parameters can more easily be driven.

To estimate β, Cox (1972, 1975) introduced the partial likelihood function, which eliminates the unknown baseline hazard λ_0 (t) and accounts for censored survival times and the frequent presence of time-dependent covariates. Time-dependence of a covariate is another important distinguishable property of survival data. A predictor variable is said to be time-dependent if its value for any given individual variable can change over time, and hence such a covariate, say Z, cannot be represented by a single value Z_i for unit i, but takes values that may change with time. For estimation of data with such distinct feature, Cox models do not use maximum likelihood

estimation but rather a maximum partial likelihood estimation method that requires only the order of the failure times, not the intervals between failure times. The basic idea for this is that under proportional hazards assumption, information about β can be obtained from the relative orderings (that is ranks) of the survival times, rather than the actual values.

The basis of the argument used in the construction of a likelihood function for the proportional hazards model is that intervals between successive death times convey no information about the effect of predictor variables on the hazards of death. This is because the baseline hazards function has an arbitrary form, so it is conceivable that $\lambda_0(t)$, and hence $\lambda(t;Z)$, is zero in the time intervals in which there are no death. This in turn means that these intervals give no information about the values of the components of β. From this argument it follows that we need only consider time order at which failures occur. To estimate the parameters of the Cox's proportional hazards model we proceed as follows:

Let t_i be the i^{th} ordered event time in the n observations, $i=1,\ldots,k$, and R_i denote the risk set at time t_i that is, the set of individuals still under study just prior to time t_i who have not yet experienced an event, but any of whom may be found to fail at time t_i, and assume that only one event occurs at any one time (that is for simplicity we assume that there are no ties in the failure times). Individuals who have previously failed or who have been censored are therefore excluded from R_i.

Let Z_i be the value of the covariate vector for the failed individual. For every individual j in R_i, the probability that an individual with covariate z_i is the one to experience the event at this time, given that one individual exist at time t_i is the conditional probability defined as follow:

P {individual i has event at t_i given that one event at t_i} =

$$P\left\{ \textit{individual } i \textit{ has event at } t_i \textit{ one event at } t_i \,/\, \textit{one event at } t_i \textit{ survival to } t_i \right\} \qquad (6)$$

From the Bayes theorem of probability (Mood, et al., 1974) where we have a partition of the sample space (death times in our case) by some events (one of the observed set of death times t_1, . . . , t_k with the corresponding risk set $R_1,\ldots,R_k$) we define the conditional probability based on this partitioned sample space as follow. If any individual is to be selected from this sample space, the probability that a randomly selected individual in the risk set of death time can experience an event/death given that any individual who survived to a particular time is selected is the probability of the individual to be included in the risk set divided by the total probability in the risk set.

Based on this argument the numerator of the above expression is simply the hazard of death at time t_i for the individuals whose vector of predictor variates is Z_i. If it is the i^{th} individuals who dies at t_i, this hazards function can be written as $\lambda_i(t;Z_i)$. The denominator is the sum of the hazards of death at time t_i over all individuals who are at risk of death at this time (the total probability theorem). This is the sum of the values $\lambda_j(t;z_j)$ over those individuals indexed by j in the risk set R_i at time t_i. Consequently, the conditional probability in (6) becomes

$$\frac{\lambda_i(t;Z_i)}{\sum_{j\varepsilon R_i}\lambda_j(t;Z_j)} \tag{7}$$

And under the proportional hazards assumption on using equation (1) we have

$$\frac{\lambda_o(t)\ \exp(\beta' z_i)^i}{\sum_{j\varepsilon R_i}\lambda_o(t)\exp(\beta' z_j)^j} \tag{8}$$

This conditional probability shows the contribution to the likelihood at each death time t_i by the individuals with covariate Z_i in the risk set R_i.

The conditioning eliminates the baseline hazards function in the nominator and dominator of equation (8). Consequently equation (8) becomes

$$\frac{\exp(\beta' z_i)}{\sum_{j \varepsilon R_i} \exp(\beta' z_j)} \qquad (9)$$

Thus the partial likelihood for β, which does not involve the base line hazard function $\lambda_0(t)$, is the product over all the failure times t_i for $i = 1, \ldots, k$ of the conditional probability (9) to give

$$PL = \prod_{i=1}^{k} \left\{ \frac{\exp(\beta' z_i)}{\sum_{j \varepsilon R_i} \exp(\beta' z_j)} \right\} \qquad (10)$$

where PL stands for the partial likelihood.

Here the product is taken over all the event time t_i and the summation is over all the risk set R_i (McCullagh and Nelder, 1983). The actual event times t_i do not appear in this expression: they are being discarded on the grounds that they tell us little about β, since the Cox model relies on the order rather than the actual values.

Since maximization of the likelihood is equivalent to maximization of the log-likelihood, we can take the partial log likelihood to obtain

$$\text{Ln (PL)} = \sum_{i=1}^{k} \left\{ \beta' z_i - Ln \sum_{j \varepsilon R_i} \exp(\beta' z_j) \right\} \qquad (11)$$

The partial likelihood derived above is valid when there are no ties in the data set, that is, no two subjects have the same event time. The proportional hazards model for survival data assumes that the hazard function is continuous, and under this assumption, tied survival times are not possible. Hence the partial likelihood above provide good estimate for the model parameters given in (1). But in most real situations the survival times are usually recorded to the nearest day, month or year, and so tied survival times are more likely to occur as the result of this rounding process. In addition to the possibility of more than one death at a time, there might also be one

or more censored observations at a time of death. Because Cox methods rely on order of events (failure times), handling those ties results in a computational problem.

To handle this real-world fact that tied time can exist, partial likelihood algorithms have been adopted to handle ties. This is to say that in order to accommodate tied observations, the likelihood functions in equation (10) has to be modified in some way. The appropriate likelihood function in the presence of tied observations had been given by Kalbfleisch and Prentice (1980). To define this appropriate likelihood function we have to first define important notations to be used in the likelihood function.

Let s_i be the vector of sums of each of the p covariate for those individuals who die at the ith death time, t_i, that is the sum of the covariates associated with the failures di at t_i, for $i= 1,...,k$.

If there are d_i deaths at t_i , the h^{th} elements of s_i is $s_{hi}=\sum_{r=1}^{di} Z_{hir}$, where Z_{hir} is the value of the h^{th} predictor variable, $h=1,\ldots,p$, for the r^{th} value of the d_i death events , $r= 1,..d_i$ who died at time t_i $i=1,..,k$.

With this notation the likelihood function in (10) is given as (Thompson et al, 2003)

$$PL = \prod_{i=1}^{k} \frac{\exp(\beta' s_i)^i}{\left[\sum_{l \varepsilon R_i} \exp(\beta' z_l)\right]^{d_i}} \qquad (12)$$

If $d_i \geq 1$ individuals exist at the same time t_i, in the presence of ties in the event times, by similar argument as for the case with no ties above in equation (11), the partial log- likelihood of (12) is given as

$$\text{Ln (PL)} = \sum_{i=1}^{k} \left\{ \beta' \sum_{i \varepsilon t_i} S_i - d_i Ln \sum_{i \varepsilon R_i} \exp(\beta' z_l) \right\} \qquad (13)$$

Equation (13) is a good approximation in the log partial likelihood of (11) if the d_i deaths at time t_i are considered to be distinct and to occur sequentially (Breslow; 1974). It is argued that the likelihood in equation (12) is the most frequently occurring partial likelihood estimate of the Cox model (1), and is an adequate approximation to estimation of the parameters in the Cox regression model when the number of the tied observations at any one time is not too large. As a rule of thumb, there should be 5% or fewer tied observations in the dataset to still assume insignificant bias in addition to the complexity of the likelihood function.

In both cases, whether the death time is tied or not, the derivation of the partial likelihood shows that actual values are discarded which normally would have to be in the full likelihood for censored observations from the Cox proportional hazards model. Thus, the likelihood function that has been obtained is not a true likelihood since it does not make direct use of the actual censored and uncensored survival times. That is to mean that, the partial likelihood is not likelihood in the usual sense in that the general construction does not give a result that is proportional to the conditional probability of any observed event. For this reason it is referred to as a partial likelihood function (Cox, 1972).

Although this may cause some possible loss of information about the regression coefficients $\beta_1,\ldots,\beta_p$, Muller (2001) had shown that the partial and full likelihood have similar features. The loss of information is not negligible for small data sets and for informative censoring. It is usually assumed that censored observations do not contribute additional information to the estimation. Otherwise, censoring is informative and estimation by partial likelihood is biased.

However, if the data set is large enough and censoring is informative, estimation of proportional hazards regression by partial likelihood method does not provide as such a significant bias in estimation. The argument is forwarded by noting that, since $\lambda_0(t)$ is completely unspecified, no additional information about β is obtained from the observation that no failure occurs (by the time intervals in which no failures occur because the component $\lambda_0(t)$ might conceivably be identically zero in such intervals). Muller (2001) showed that there is little loss in assuming this,

since censored observations enter the risk set at each observation but do not contribute to the numerator of the partial likelihood, thus PL above can be treated as if it were a likelihood.

Nonetheless, it has been shown informally by Cox that the method used to construct this likelihood gives maximum "partial" likelihood estimates that are consistent and asymptotically normally distributed with asymptotic covariance matrix of second partial derivatives of the log likelihood function. It then follows that although the resulting estimates are not as efficient as maximum-likelihood estimates for a correctly specified parametric hazard regression model, the same asymptotic results for β hold for estimation from partial likelihood as for the usual likelihood function.

As a usual technique of estimation of the parameters by the maximum likelihood method, the parameters in the model can be estimated by equating the derivative of the log-likelihood of equation (11) or (13) to zero depending on whether absence or presence of the tied death times for an individuals. Then the maximal condition for an estimate can be checked from their corresponding second derivatives (information matrix). If the determinant of the information matrix obtained is positive then the estimate is a maximum likelihood estimates otherwise it is not (Mood, et al., 1974).

The most communally used statistical software for analysis of survival data, which we are going to employ in our analysis (estimation of Cox proportional hazards model parameters) as well for the test of proportionality assumption are SAS, STATA, and SPSS.

3.3.3 Assessing the Goodness of fit of the Model

Like any other statistical model, the validity of Cox regression analysis is based on the plausibility of the assumption made regarding this model. Thus in order to use the Cox model, we must check the assumption of whether the effects of covariates on risk remain constant over time. In fact, this proportionality assumption is a critical assumption of proportional hazard model and must have to be checked for each covariate. The proportional hazards assumption

may not hold in some of the several studies. Predictors of interest are often measured only at time zero, even if their values may change later, so that a time-dependent covariate must be treated as a fixed one. Further, the impact of a predictor, which is by definition fixed in time, may change during time. This shows that it is common for a covariate to fail to satisfy the assumption of proportional hazards. Hence the assumption of the proportional hazards model is questionable.

With the growing popularity of the semi parametric proportional hazards of Cox (1972), it has become increasingly important to find a convenient way to detect when such models are poorly specified. A number of goodness-of-fit test for the Cox model have been proposed. Some of these are indicated in Schoenfeld (1980), Andersen (1982), Moreav, O'Quigley and Mesbah (1985), and Moreav, O' Quigley and Lellouch (1986) all focusing on the two basic explicit assumptions of the Cox proportional hazards model.

The need for the development of different methods for the test of the proportional hazards model to make any inference from the data result is that: Estimation of proportional hazards models when in fact hazards are non-proportional results in coefficient's biase estimate, incorrect standard errors, faulty inferences about the substantive impact of covariate variables and thereby decreased power of significance tests. In particular if a covariate fails the proportionality assumption of proportional hazards model then for a hazard ratio that increases over time for that covariate, relative risk is overestimated (that is, for diverging hazards, coefficient estimates are inflated). For ratios that decrease over time, relative risk is often underestimated (that is, for converging hazards, coefficient estimates are deflated and biased toward zero). Correspondingly, standard errors are incorrect and the power of significance tests decreases (Box-Steffensmeier and Zorn, 2001).

Because of this, as different researchers use the proportional hazards models, in their research study especially in medical areas (or in psychiatry like the one we are focusing on in this study), care must be taken to examine the extent to which the assumption of the proportional hazards model is consistent with the data.

Studies have suggested that several graphical techniques can be used to assess model assumptions and the fit of a potential Cox model. The fit of a Cox model can be assessed graphically using martingale residuals or partial residuals.

The martingale residual is the difference between the observed and the expected numbers of events for a given individual. These residual measures the excess number of events that have occurred compared with the expected number. The estimated cumulative hazard at the time can be viewed as the expected number of events for the individual in some time intervals. To assess the proportionality of the proportional hazards model we would plot the martingale residuals on the y-axis against a linear predictor (DeBeta(s)) of each covariate on the x-axis. DeBeta(s) is a measure of how much a given case will affect the regression for a given covariate. Under the assumption of proportionality of the proportional hazards model, there is no pattern of correlation in the plot.

In the Cox framework, we assume a log linear functional form for the covariates. However, covariates in the Cox model can have non-linear functional forms just as in the linear model. For continuous covariates, this assumption is often violated as the effect is highly nonlinear. Assuming a log-linear functional form when the nonlinear form is correct can cause specification errors leading to erroneous statistical conclusions (LeBlanc and Crowley, 1999). Thus the functional form of the model has to be assessed. The martingale residual is usually plotted to assess for the functional form of the covariates in the model. It can be done by plotting the martingale residuals on Y axis against covariate on the X axis. A plot of martingale residuals for each covariate may reveal departures from the linearity of the functional form assumptions (Therneau, Grambsch, and Fleming, 1990) if its plot forms a linear trend.

The partial residual which is also called Schoenfeld residual is a measure of the difference between the covariate for the i^{th} individual and a weighted average of the covariate over the risk set at the time the i^{th} individual event (Schoenfeld, 1982). This is a plot of scaled Schoenfeld residuals on the y-axis against survival time on the x-axis, with such plot per covariate. The partial residual methods are the most common and preferred methods for testing non-proportionality in Cox models. In a well fitted model, distribution of residuals over time is

random (do not show a particular trend). Otherwise a plot of partial residuals for a given covariate may reveal a violation of the proportional hazards assumption.

Box-Steffensmeier and Zorn (2001) had shown that the average value of residuals at any time should be zero if the effects of the covariate being plotted are proportional. In such case Locally Weighted Smoothing Scatter plots (Lowess) smoothing line summarizing the residuals should be close to the horizontal reference line for the y-axis, since the average value of residuals at any time should be zero if the effects of the covariates being plotted are proportional.

The Grambsch-Therneau global test of non-proportionality uses partial residuals for the test of proportional hazard assumption. If the global test of Grambsch-Therneau coefficient is non-significant, at least one of the covariates is non-proportional. In order to run such test, for the j^{th} covariate, Grambsch and Therneau (1994) express a time-varying coefficient as

$$\beta_j(t) = \beta_j + \gamma_j g_j(t)$$

where $g_j(t)$ is a specific function of time and γ_j is a slope parameter. It is a form of simple linear regression model. They pointed out that, a plot of the scaled Schoenfeld residuals by the event times may assess whether the coefficient γ_j is zero or not, whatever the function $g_j(t)$ might be. A non-zero slope is evidence against proportional hazard assumption.

Another important method used for the test of proportional hazards assumption is the use of time interaction test. According to this test, if the assumption of the proportional hazards is not violated for a given numerical (continuous or categorical) covariate, then the interaction term between that covariate and time or more commonly log of time can be added to the model as in regression, and should have a regression coefficient not significantly different from zero. If the time interaction effect is significant for a covariate, then the proportional hazards assumption is violated and the covariate should be modeled as a time-dependent covariate.

For binary covariates in general, a comparison of non-parametric hazard estimates may be sufficient to decide on the proportionality hazard assumption because if the hazard were

proportional, the survival curve for the two binary conditions would separate exponentially. Hence the hazard estimates curve for the two binary conditions would not cross each other.

For continuous covariates, it is not sufficient to rely only on the stratified hazard estimates to assess proportionality hazard assumptions because the choice of stratification points is subjective and arbitrary in that case. In such a case, we rather use other alternatives to assess proportional hazards assumptions across the values of continuous covariates. One possible alternative is the use of time-varying covariates (Grambsch and Therneau, 1994). As discussed above, if the coefficient associated with a covariate is not constant over time, then the impact of that covariate on the hazard varies over time, leading to non-proportional hazards. Graphically this means that if proportional hazards assumption holds, a plot of the coefficient versus time will be a horizontal line.

All the above tests are the different techniques for the test of the assumption of proportional hazards model. As it was indicated earlier, the response variable to be included in proportional hazards model can be categorical or continuous. The category of the variables (continuous or categorical) may sometimes tell as the type of simpler and more advantageous test to be employed.

If the proportional hazards assumption does not hold for one particular covariate, it may be possible to build interactions between covariates and time into the Cox regression model in which case interactions are themselves time-dependent, or to divide the data into strata based on that covariate, and fit a stratified model.

In the next chapter (chapter four) we will estimate the model parameters, check whether the assumptions of the proportional hazards are met and provide the results of the analysis. In testing the proportionality assumption of proportional hazards model, we would use any one of the appropriate techniques discussed earlier. In testing for the log-linearity of the functional form of the model covariate we use the martingale residual test as discussed above. And finally we will make appropriate interpretations, conclusions and recommendations regarding the study results.

Chapter Four

Data Analysis

4.1 Descriptive analysis

In any data analysis it is useful to provide some descriptive results. Before proceeding to more complicated models, we would make a descriptive analysis that will use in highlight to our subsequent findings. Here we start with the test of equality across the different levels of a categorical variable. In this case we use log rank test to test the null hypothesis that there is no difference between the probabilities of an event occurring at any time point for each population.

Table 1: Log-rank test for equality of the different levels of a covariates

covariates	Chi-Square	df	Sig.
gender	17.665	1	.000
Marital status	32.401	4	.000
Sub use	49.954	2	.000
Phy illness	3.686	3	.297
Dur illness	46.780	3	.000

The table shows that the different levels of gender, marital status, substance use and duration of illness are statistically not the same in experiencing the death event. The log-rank, non-parametric test results suggest that gender, marital status, substance use and duration of illness are significant covariates whose different levels have an impact in the survival longevity of schizophrenic patients while physical illness does not have an impact.

Based on the results of the log-rank test we would consider the levels of the marital status, Duration of illness, Substance use and Gender as different factors and assume that each of these has its own effect on the survival longevity of schizophrenic patients.

4.2 Univariate analysis

For the continuous variables we will use a univariate Cox proportional hazards model analysis in order to have an idea about each of the variables. If the predictor is highly significant in the univariate Cox proportional model where only one predictor is included in the model then it is advisable to include this variable to the model. The continuous covariates in this study are Age1 to be defined for age at onset of the illness, Age2 to be defined as the current age of schizophrenic patients and Edu is to be defined as the educational levels of schizophrenic patients measured in years

Table 2: Univariate analysis for the covariates

Variable name	Coefficient estimate	Standard error	Wald statistic	DF	Sig
Age1	.004	.008	.225	1	.635
Age2	.001	.009	.008	1	.929
Edu	.003	.002	1.634	1	.201

The table shows that none of the covariates namely age at the onset of the illness, levels of education (in years), and current age of the schizophrenic patients separately has a significant effect on the survival longevity of patients. It points out that we can remove age at onset of the illness, current age of schizophrenic patients and level of education (in years) from our final model. Although the univariate Cox proportional hazards model results indicate the non-significant effect of the covariates on the survival longevity, in our study we will include all the covariates to our model. Some variables which have no impact on the survival time may have a significant effect when used together with other covariates in the model. As an alternative it may

be used as a confirmation in the Cox proportional hazards analysis whether or not the variable is not important in the determination of the effects.

In general, the log rank test we discussed in the descriptive analysis is used to assess whether there is a difference between the survival times of different levels of covariate that are assumed to affect survival longevity of schizophrenic patients. Whereas the univariate Cox's proportional hazards model identify the effect of a covariate on the survival longevity of schizophrenic patients. Both the log rank test and the univariate Cox proportional hazards method do not allow other predictor variables to be taken into account. On the contrary of this Cox's proportional hazards model is analogous to a multiple regression model and enables us to analyze the effect of multiple covariates on the survival longevity of schizophrenic patients.

4.3 Estimation of model parameters

The model parameters to be estimated here in the Cox proportional hazards models are the coefficients β_j of the covariates in equation (1). To estimate those model parameters, we basically make use of the estimation argument we discussed in chapter three with declaration of the covariate variables.

- Time: Month of first experience of death in the five years follow up.
- Status: the event indicator, equal to 1 for those that have experienced an event in the five years follow up and 0 for those who have not.
- Marital: Marital status is a categorical variable, with code 1(Married), 2 (Never married), 3(Separated), 4(Divorced), or 5(Widowed).
- Sub use: Substance use is a Categorical variable, with code 0(not at all), 1(a little), or 2(Some)
- Age1: Age in years at the onset of the illness.
- Age2: Current age of schizophrenic patients.
- Gender: a dummy variable coded 1 if the individual is male and 2 if female.
- Edu: levels of education in years.

- Dur illness: Months in the hospital due to psychiatric problem (duration of illness) coded as 0(none of the periods), 1(about one month or less), 2(about two months), or 3(three or more months).
- Phy illness: physical illness or disability is a categorical variable coded as 0(not at all), 1(a little), 2(some) or 3 (a lot).

With these declarations of the variables in the Cox regression model, different statistical software can provide us the estimate of the parameters. For this purpose we use SPSS to estimate the parameters. We analyze eight predictor variables related to the hazard of death of schizophrenia patients and a response variable time-to-event. The results are given in Table 3 below.

Table 3: Partial Likelihood Estimates for Fitted Proportional Hazards Models

Variable In the model	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Edu	.001	.002	.108	1	.743	1.001	.996	1.005
age1	.006	.009	.500	1	.480	1.006	.989	1.025
Phy illness			.281	3	.963			
Phy illness(1)	8.625	75.130	.013	1	.909	5571.474	.000	4.979E+067
Phy illness(2)	8.373	75.132	.012	1	.911	4330.629	.000	3.884E+067
Phy illness(3)	8.687	75.131	.013	1	.908	5924.749	.000	5.302E+067
Dur illness			24.801	3	.000			
Dur illness (1)	-.725	.188	14.826	1	.000	.484	.335	.700
Dur illness (2)	-.982	.741	1.756	1	.185	.374	.088	1.601
Dur illness (3)	.123	.226	.294	1	.587	1.130	.726	1.760
Marital			17.414	4	.002			
Marital (1)	-.627	.433	2.094	1	.148	.534	.229	1.249
Marital (2)	-.727	.461	2.486	1	.115	.484	.196	1.193
Marital(3)	-.197	.482	.167	1	.683	.821	.319	2.111
Marital(4)	.094	.431	.048	1	.827	1.099	.472	2.560

Gender	-.290	.167	3.034	1	.082	.748	.540	1.037
age2	-.014	.011	1.712	1	.191	.986	.965	1.007
Sub use			26.085	2	.000			
Sub use (1)	-1.002	.207	23.437	1	.000	.367	.245	.551
Sub use (2)	-.456	.257	3.136	1	.077	.634	.383	1.050

Table 3 shows a computer output showing the results of the fitted hazards model: the estimate of the slope parameter for each covariate, standard errors of these estimate, ratio of each parameters estimated to its standard error (Wald statistic) which is assumed to have standard normal distribution under the null hypothesis and the p-value based on the values of the Wald statistic for testing the null hypothesis that the parameter is zero in the population. The value of the Wald statistic and p-values given in Table 3 are useful to assess the significance of each covariate in the model. The table also shows the estimate of the hazards ratio and the 95% confidence intervals for the estimated hazards ratio which can be calculated as [exp (B – 1.96 S.E (B)), exp (B +1.96 S.E (B))].

Based on the results of Table 3 we look for predictors having a statistically significant relationship with the hazards (death of schizophrenic patient) often focusing primarily on p-values. Setting a suitable α-level, we consider the predictor to be significant if its p-value is smaller than the target value. Now for the covariate which is assumed to be statistically significant based on the above test, we would make a meaningful interpretation in the following subsection.

4.4 Test for the significance of model parameters

Statistical models are of little use unless an investigator can test for the significance of the model parameters, interpret and present them in clear and persuasive ways. The test for the significance of the model parameters can be made by testing the overall fitted model with the likelihood ratio test, Wald test, or score chi-square test. In our study we use the chi-square tests to establish if

there are the relationships between time and all the covariates in the model. To check for the overall fit of the model we make tests of the omnibus null hypothesis that all of the β's are zero. The statistical test for this hypothesis can be rewritten as

$\mathbf{H_0}: \beta = (\beta_1, .., \beta_p)' = \mathbf{0} = (0, . . ., 0)'$ versus the alternative

$\mathbf{H_1}$: $\beta_j \neq \mathbf{0}$ for at least one j = 1, . . ., p at $\alpha = 0.05$.

If at least one of the β's is non-zero we reject the null hypothesis stated under $\mathbf{H_0}$ and conclude that at least one of the covariate in the model has an effect on the survival longevity of schizophrenic patients. If the model is significant for at least one of the covariate coefficient then we can also make further test for each covariate. In this case statistical test for individual significance of the parameters can be rewritten as

$\mathbf{H_0}: \beta_j = 0$ versus the alternative

$\mathbf{H_1}$: $\beta_j \neq 0$ at $\alpha = 0.05$ for $j = 1, . . ., p$.

Finally, we would make an investigation on the assumptions of the model concerning the important covariates.

Table 4: Tests of Model adequacy and Model Coefficients

Omnibus Tests of Model Coefficients

-2 Log Likelihood
1904.230

Omnibus Tests of Model Coefficients (a)

-2 LL	Overall (score) or Global			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1806.259	113.178	16	.000	97.971	16	.000	97.971	16	.000

The likelihood ratio test of the model, also called the Omnibus test or -2log likelihood (-2LL), tests whether the model as a whole is significant. In Table 4 the overall significance is less than 0.05 indicating at least one of the variables in the model at this stage is significant. The table also shows that the change from the previous step of significance is less than 0.05 indicating that the variable added in that step is significant. The same thing is true for a change from the previous block. In fact since there is no block change from previous step is same as change from previous block.

In Table 4 above, the null (intercept-only) model has -2LL = 1904.230. The full model has -2LL = 1806.254 with the model chi-square difference of 113.178 which is significant. This shows that at least one of the covariates contributes significantly to give explanation about the duration in months to event of schizophrenic patients. It also means that the model is significantly better than the null model, which is the time-only model when all covariates are zero. Thus, we can conclude that the Cox proportional hazard model fit the data well.

The test for the significance of each covariate can be made either by constructing a confidence interval for each covariate coefficient or by comparing the ratio of each Cox regression coefficient to its standard error and the value of the Wald statistic. The confidence interval allows us to assess how big the effect is. Based on those results we would make identification of statistically significant covariates, which is assumed to contribute to hazards rate for schizophrenic patients.

In medical area we use proportional hazards model because we are interested to know which covariate has an effect on the hazards rate and to what extent. Thus, the covariates that are tested to be statistically significant.

In identifying a covariate that have a statistically significant effect on the survival longevity of schizophrenic patient we use the P-value which is based on the Wald statistic. If the p-value is greater than 0.05 (the usual standard α level), then the covariate effect cannot be assumed to be

different from zero. On the other hand if the p-value is less than or equal to 0.05 then we conclude that the variable contributes to the model.

The covariates which are assumed to have significant effects on the hazards of schizophrenia patients are: Marital status, Duration of the illness categorized as none of the periods, about one month or less, about 2 months and three or more months, and Substance use.

The coefficients $\beta_1,...,\beta_p$ estimated by Cox regression can be interpreted in a similar manner like that multiple logistic regression. The meaning of β_j in the Cox regression model is that suppose we have two groups: Group1 with covariates $Z_1 \ldots Z_j \ldots Z_p$ and Group2 with covariates $Z_1 \ldots Z_{j+1} \ldots Z_p$, then change in hazard rate on log scale is given as β_j which is equal to the log hazard rate of Group 2 minus Group 1.

The sign of the estimated regression coefficient β_j indicate that for $\beta_j > 0$ increase in the covariate say Z_j keeping all other covariate at the same level for this individual, result in the increment of the hazard rate (an increased probability for event to occur, death of schizophrenic patients in our case) which consequently indicates a low survival longevity due to those risk factors. The opposite situation can be said for the case $\beta_j < 0$.

Sometimes we may be interested in the relative risk or hazard ratios of the above two groups. Under such case we are estimating $\exp(\beta_j)$ which is the ratio of the hazard rate in Group 2 to that of the hazard rate in Group 1. The hazard rate in Group1 is a multiple of the hazard rate in Group2. The meaning of the estimated hazard ratio is, for $\exp(\beta_j) > 1$ increase in the covariate Z_j implies an increment in the hazard rate by a factor of $\exp(\beta_j)$ and a decrease in the survival longevity of schizophrenic patients due to this risk factor. The converse is true for $\exp(\beta_j) < 1$.

Hazard or odds ratio greater than one are associated with increased hazards (few months for events to occur) of the event while hazards or odds ratio less than one are associated with decreased (more months for events to occur) hazard of the event. In other words, an odds ratio

greater than one correspond to the positive Cox regression coefficient, while a value less than one corresponds to negative Cox regression coefficient.

In both cases the exponentiated coefficient of the output are interpretable as multiplicative effects on the hazard. This algebraic interpretation is useful only if the proportional hazards assumption is met. If the effects of predictors differ over time, the risk factor is no longer parallel and so it makes little sense to talk about one factor being rescaled to generate the other. Thus the interpretation becomes misleading. Under such cases we rather use a graphical interpretation for our analysis, fitted survival plot which demonstrate when this effect rises, or fall, or remain constant with the passage of time or to use the stratified estimate to make a valid interpretation.

4.5 Tests for proportional hazards

As with the other regression models we check for departures from assumption such as that of proportional hazards and the required functional form in the Cox model.

In this sub-section we will check if the proportionality assumption of the Cox regression model holds. As it was described in Chapter three graphical diagnoses of scaled Schoenfeld residuals supported by chi-square statistic can be employed to assess the proportional hazard assumption. The scaled Schoenfeld graphical method assesses the proportionality of the proportional hazards for each covariate.

In this graphical test conventionally, we use the x-axis as the survival time and the y-axis as the partial residuals for each category of a given covariate. Under the assumption of proportionality of the proportional hazards model, the distribution of residuals over time is random. The randomness can be detected by adding a loess smoothing line which can be either a horizontal line or any straight line. In this study our reference line is taken as a horizontal line.

Graphical assessment of the proportional hazards assumption

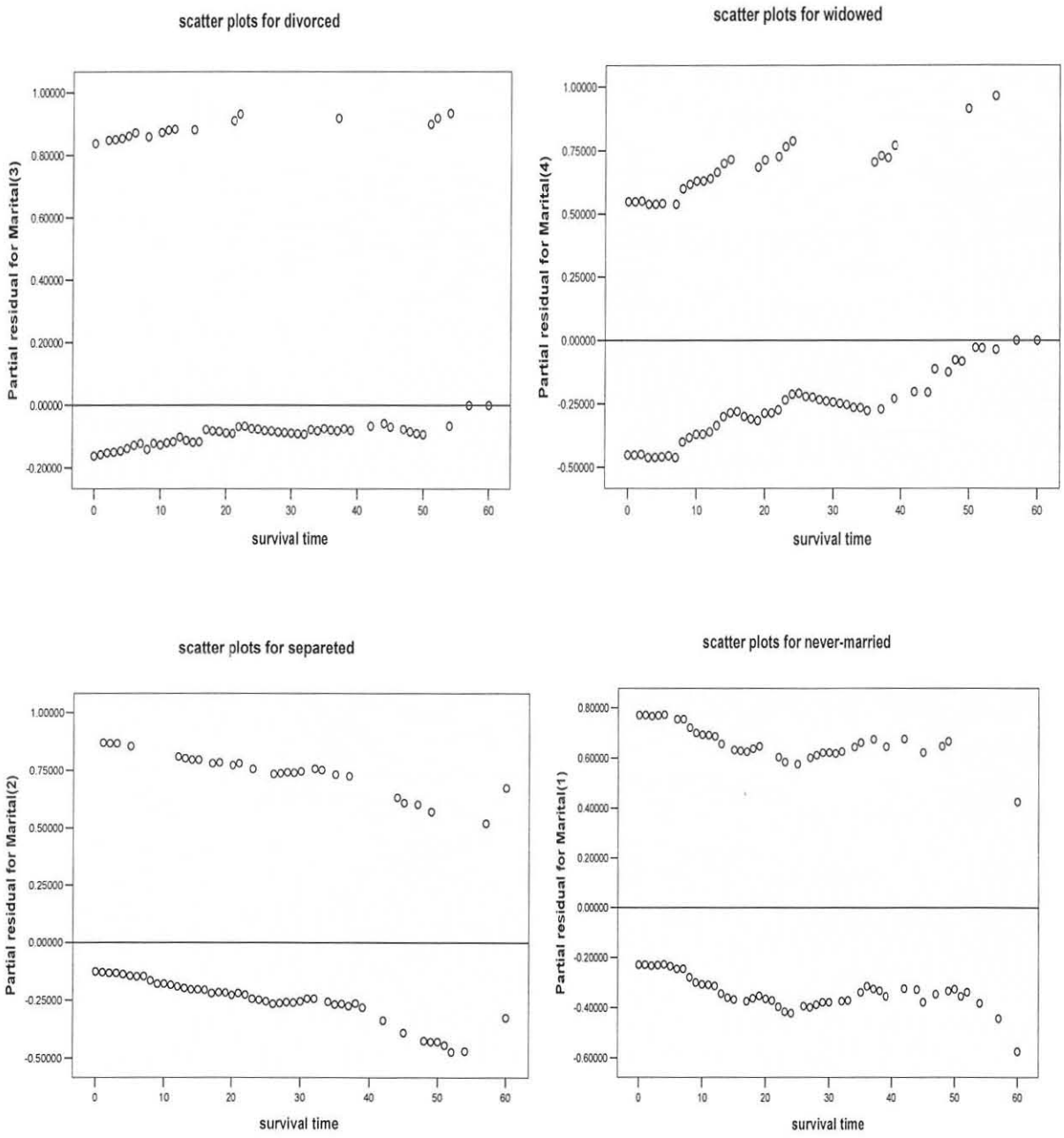


Fig 1: Graphical proportionality test for marital status

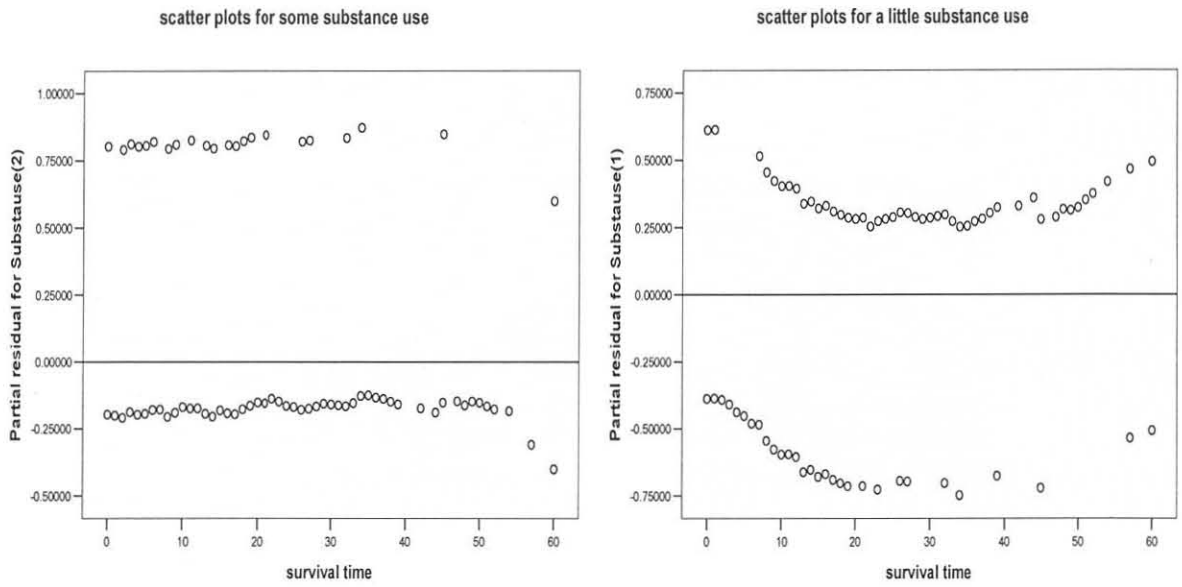
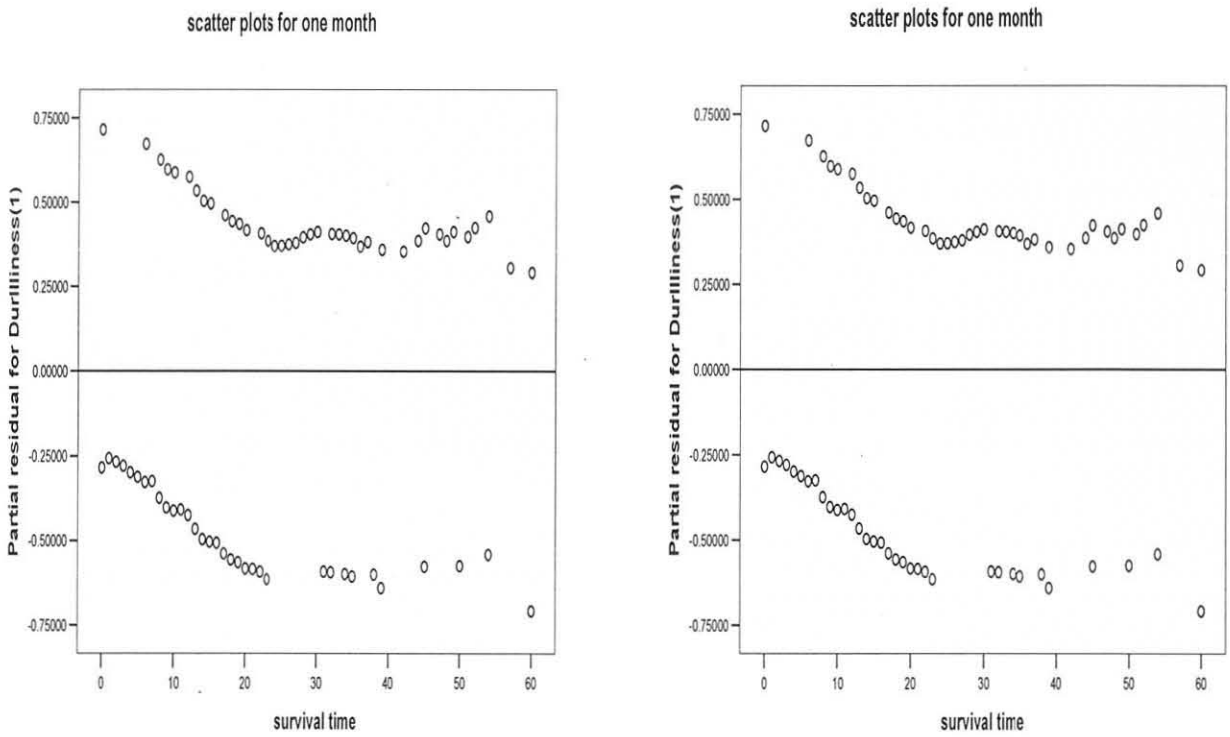


Fig 2: Graphical proportionality test for substance use



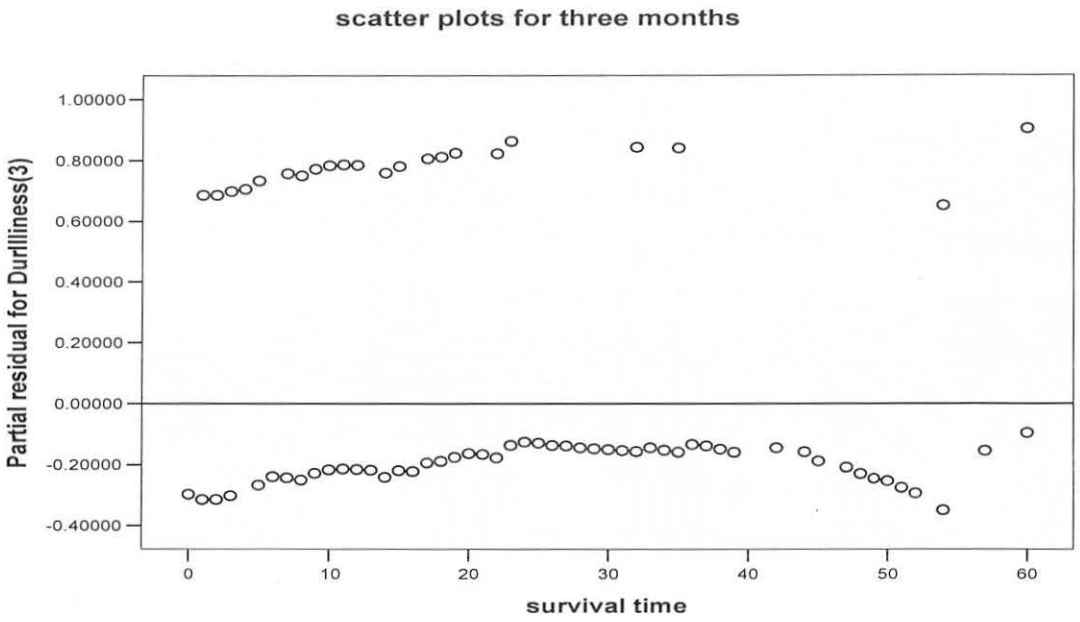


Fig 3: Graphical proportionality test for Duration of illness

The graphical displays show plots of the partial residuals against the survival time for each case of the covariate marital status, substance use and duration of illness. In Fig 1 the marital status of schizophrenic patient does not manifest randomness in the residual over time. Each of the categories of the marital status shows trends below and above the horizontal reference zero line. On the other hand the average value of the residual at any time is not zero. Differently speaking the residuals are not at the same vertical distance below and above the zero reference line on average. In Figures 2 and 3, respectively for substance use and duration of illness of schizophrenic patients the plots show similar trends. Thus the plots show the failure of the proportionality of the proportional hazards model.

The graphical test is not enough to be certain of the proportionality of the proportional hazards model. The reason is that we cannot quantify the magnitude of the deviation from the horizontal reference line. By looking at the graph different persons may give different interpretations. This is to say that the graphical method of assessing the proportionality of the proportional hazards model is more or less subjective but we use it as a supportive argument for proportionality test.

The graphical test for the proportionality of the proportional hazards model argument in the figures considered earlier can be checked further by statistical tests. One of the statistical tests for proportional hazard model is to generate the time-dependent covariates by creating interactions of the predictors and a function of survival times (usually covariate times log of time) and including these in the model. If any of the time-dependent covariates are significant then those predictors do not show a proportional effect over the study period. That is the proportional hazard assumption fails to hold. In performing the statistical test the SAS program gives the following overall tests for the proportionality effects of the covariates.

Table 5: Linear Hypothesis Testing Results for proportionality

label	Wald Chi-square	DF	Pr >chi-square
Test-proportionality	9.1556	3	0.0273

In Table 5 the Wald chi-square value is greater than the tabulated value of a chi-square distribution with three degree of freedom at 0.05 levels of significance. This shows that the time-dependent variable is collectively significant. So we do have enough evidence to reject proportionality and will assume that we are not yet satisfied with the assumption of proportionality of this model. This in turn shows that at least one of the covariates is a time-dependent covariate showing rejection of the proportionality assumption of the proportional hazards model.

Having rejected proportionality, one way of accommodating non-proportional hazards is to divide the data into strata based on the value of that time-dependent covariate. The only change in the model is the addition of strata statement together with the time-dependent covariate in the SAS program and exclusion of that particular variable (marital status, substance use or duration of illness whenever it is time-dependent) from the model. The assumption is that we are fitting separate models for each level of a particular time-dependent covariate under the constraint that the coefficients are equal but the baseline hazard function are not equal. This is to say that each

stratum is permitted to have a different baseline hazards function, while the coefficients of the remaining covariate are assumed to be constant across strata.

However, the remedies to the failure of the proportional hazards model assumption made through stratification have a disadvantage. Especially for the continuous covariate, categorizing of this variable into some levels may be subjective. Thus an alternative way is to build interactions between covariates and time in the Cox regression model.

Here in our study we consider interactions between covariates and time, and then include the interaction variable to the model. The estimate of the resulting model is given in Table 6 and 7.

Table 6: **Omnibus Tests of Model Coefficients of the null model**

-2 Log Likelihood
1791.922

Table 7: **Omnibus Tests of Model Coefficients of the full model**

-2 LL	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1376.745	629.211	12	.000	415.177	12	.000	415.177	12	.000

The Omnibus test of model coefficient in Table 6 indicates that the null (intercept-only) model has -2LL which is equal to 1791.922. The full model shown in Table 7 has the -2LL which is

equal to 1376.745, a model chi-square difference of 629.211, which is significant at the given level of significance. This shows that the model with the included variables explains the survival duration better compared to the null model. On the other hand this means that at least one of the variables included in our model has a significant contribution in explaining the survival longevity of schizophrenic patients. To make an identification of this variable and make further statistical interpretation we need to estimate the model parameters.

Table 8: Partial Likelihood Estimates for Fitted Proportional Hazards Models

Variable in the model	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
durt	-2.162	.324	44.639	1	.000	.115	.061	.217
murt	-1.349	.215	39.314	1	.000	.260	.170	.396
Sub use			17.593	2	.000			
Sub use (1)	-3.195	.765	17.432	1	.000	.041	.009	.184
Sub use (2)	-1.186	.384	9.527	1	.002	.305	.144	.649
Marital			46.022	4	.000			
Marital(1)	-7.414	1.150	41.563	1	.000	.001	.000	.006
Marital(2)	-5.325	.854	38.847	1	.000	.005	.001	.026
Marital(3)	-3.240	.640	25.652	1	.000	.039	.011	.137
Marital(4)	-1.583	.493	10.311	1	.001	.205	.078	.540
Dur illness			51.586	3	.000			
Dur illness (1)	-9.475	1.332	50.623	1	.000	.000	.000	.001
Dur illness (2)	-6.443	1.080	35.589	1	.000	.002	.000	.013
Dur illness (3)	-2.294	.380	36.436	1	.000	.101	.048	.212
Sub uset	-1.215	.326	13.925	1	.000	.297	.157	.562

The estimation results of Table 8 also show the supportive argument of the results in Table 6 and 7. In Table 8 one can conclude that the covariates duration of illness, marital status, substance

use and the interaction variables duration of illness interacted with log of time (durt), marital status interacted with log of time(mart), and substance abuse interacted with log of time(sub uset) included to the model are all statistically significant at 0.05 level of significance. Here the generated covariate with the interaction of the covariate and the log of time is significant. The significance of this covariate supports that the proportional hazards assumption fails to hold as indicated in Table 5.

The significance of the interaction variables (duration of illness interacted with log of time, marital status interacted with log of time, and substance abuse interacted with log of time) necessitate the need to include the interaction variable to the model. Moreover Table 8 shows that marital status, duration of illness and substance use are time-dependent covariates. The coefficient of interaction covariate is negative for each of these covariates implying that all of these variables reduce the hazards ratio. When we come to the detail of each covariate, for example for marital status interaction with log of time estimated value of the coefficient β is -1.349. Exp($\hat{\beta}$)= exp(-1.349) is 0.260 with a 95% confidence interval of 0.170 to 0.396, meaning that the hazards ratio decreased by a factor of 0.260 due to the interaction of marital status and log time effects. In a similar way the hazards ratio decreased by a factor of 0.115 and 0.279, respectively, due to the interaction effect of duration of illness and log of time and substance use and log of time.

Since the variable in Table 8 is time-dependent, we use a graphical interpretation. The graph in the appendix shows a decreasing hazards rate with time. In the graph the patients with the lower curve indicate higher death rate than the upper curve. When we come to the specific covariate for example, for marital status the curves are in the increasing order for Divorced, Separated, Widowed, Married and Never Married respectively indicating that keeping assuming the same level for all other covariate Divorced schizophrenic patient have the highest death rate while Never married schizophrenic patient have the lowest. In similar way the graph for substance use and duration of illness are interpretable.

Another way to accommodate the time-dependent covariate and to make a valid interpretation from the resulting estimate is to stratify the data set based on the value of the time-dependent covariate.

Table 9: Partial Likelihood Estimates for Fitted Proportional Hazards Models

(Stratification by substance use)

variable	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
DurIlliness			27.549	3	.000			
DurIlliness(1)	-.777	.181	18.346	1	.000	.460	.322	.656
DurIlliness(2)	-1.150	.727	2.506	1	.113	.317	.076	1.315
DurIlliness(3)	.084	.217	.149	1	.699	1.088	.711	1.664
Marital			13.055	4	.011			
Marital(1)	-.526	.406	1.683	1	.195	.591	.267	1.308
Marital(2)	-.529	.413	1.644	1	.200	.589	.262	1.323
Marital(3)	-.182	.456	.159	1	.690	.834	.341	2.038
Marital(4)	.100	.411	.059	1	.808	1.105	.494	2.472

For categorical covariates in Table 9, one must interpret the results with respect to the reference category. In our analysis we use the first category (married, none of the period) as the reference category. For example, for the marital status of psychiatric patients, the estimated covariate coefficient $\hat{\beta}$ = -0.526 for the risk factor NEVER MARRIED implies that the hazards ratio is exp ($\hat{\beta}$) = 0.591 on the average. This shows that the estimated odds that NEVER MARRIED psychiatric patients with severe mental illness die is 0.591 times greater than for MARRIED. Differently speaking the quantity exp ($\hat{\beta}$) = 0.591 indicates the instantaneous relative risk of an event (death in this case), at any time, for an individual with the risk factor NEVER MARRIED compared with an individual with risk factor MARRIED, assuming that the two individuals do not differ from one another except in their marital status.

The estimated covariate coefficient $\hat{\beta}$ = -0.529 for the risk factor SEPARATED implies that the hazards ratio is exp ($\hat{\beta}$) = 0.589 on the average. This shows that the estimated odds that SEPARATED schizophrenic patient dies is 0.589 times greater than for MARRIED. The

estimated covariate coefficient $\hat{\beta}$ = -0.182 for the risk factor DIVORCED implies that the hazards ratio exp ($\hat{\beta}$) = 0.834 on the average. This shows that the estimated odds that DIVORCED schizophrenic patient dies is 0.834 times greater than for MARRIED. Similarly, the estimated covariate coefficient $\hat{\beta}$ = 0.100 for the risk factors WIDOWED implies that the hazards ratio exp ($\hat{\beta}$) =1.105 on the average. This shows that the estimated odds that WIDOWED schizophrenic patient dies is 1.105 times that of MARRIED schizophrenic patients.

The sign of the estimated covariate coefficients do have an implication on the hazard rate of schizophrenic patients, for example for the covariate MARRIED with the risk factor NEVER MARRIED the value of the estimated coefficient $\hat{\beta}$ = -0.526. The negative sign implies that being NEVER MARRIED reduces the hazard rate of a schizophrenic patients and the hazard ratio value consequently implies that holding the other covariate constant, an additional month of being NEVER MARRIED reduces the hazard rate of schizophrenic patients by a factor of exp ($\hat{\beta}$) =0.591 an average as compared to the MARRIED individuals. Similarly for schizophrenic individuals at the same level of all other covariates, being SEPARATED and DIVORCED reduces the hazard rate of these schizophrenic patients by 0.589 and 0.834 on average respectively as compared to the MARRIED individuals. To the contrary of this keeping all other covariates at the same level, an additional month of being WIDOWED increases the hazard rate of schizophrenic patients by 1.105 on the average compared with the MARRIED schizophrenic patients.

Another way to interpret this scaling factor is in terms of percentage difference in risk. The general rule is the percentage difference in risk per unit difference in the predictor is given as 100(exp ($\hat{\beta}$) - 1), Allison (1995). Thus an estimate of exp ($\hat{\beta}$) or 100(exp ($\hat{\beta}$) -1) are therefore multipliers of, or percentage of increase in the odds of an event. For example in the marital status of schizophrenic patients with NEVER MARRIED status having the estimated parameter -0.526 and hazard ratio of 0.591, can be interpreted as being NEVER MARRIED decreases the odds of death for schizophrenic patients by 40.9% compared to the married schizophrenic patients

assuming that both patients are at the same level in all other factors. In a similar way we can make interpretations for other categories of the marital status.

The durations in of months a schizophrenic patient had been in the hospital due to psychiatric problem do have a significant impact on the hazard rate of schizophrenic patients. The result of the analysis shows that when we compare the schizophrenic patients who had never been in a hospital due to psychiatric problems assuming that the individuals are at the same level for all other covariates, schizophrenic patients about 2 months and about three or more months have respectively 0.460, and 0.317 times greater than patients who were never admitted into a hospital due to psychiatric problems.

The sign on the estimated covariates coefficient is negative for the risk factor less than one month and for about two months. It indicates that additional months of being in hospital due to psychiatric problem for one month or less and for about two month, reduces the hazards of schizophrenic patients by , 54%, and 68.3% respectively. It takes some more months until the occurrence of an event by schizophrenic patients due to these risk factors. And the positive sign for the covariate estimate of the risk factor three or more months indicate that an additional month of being in hospital due to psychiatric problem for three or more months increases the hazard rate of schizophrenic patients by 8.8%. It is again equivalent to saying that schizophrenic patients who had been in hospital due to psychiatric problems for three or more months have fewer months until the occurrence of an event (death) compared with the one who had not been in hospital due to psychiatric problem.

Table 10: Partial Likelihood Estimates for Fitted Proportional Hazards Models (Stratification by marital status)

Variable	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
DurIlliness			30.483	3	.000			
DurIlliness(1)	-.844	.180	21.898	1	.000	.430	.302	.612
DurIlliness(2)	-1.009	.726	1.929	1	.165	.365	.088	1.514
DurIlliness(3)	.035	.219	.026	1	.871	1.036	.675	1.590
Substause			28.840	2	.000			
Substause(1)	-1.038	.203	26.248	1	.000	.354	.238	.527
Substause(2)	-.489	.248	3.880	1	.049	.613	.377	.998

In Table 10 the stratifying variable is marital status, with the reference category to be none of the period, no substance use respectively for duration of illness and substance use. The durations in of months a schizophrenic patient had been in the hospital due to psychiatric problem do have a significant impact on the hazard rate of schizophrenic patients. The result of the analysis shows that when we compare the schizophrenic patients who had never been in a hospital due to psychiatric problems assuming that the individuals are at the same level for all other covariates, schizophrenic patients with about one month or less, about 2 months and about three or more months have respectively 0.430, 0.365, and 1.036 times greater than patients who were never admitted into a hospital due to psychiatric problems.

The sign on the estimated covariates coefficient is negative for the risk factor less than one months and 2 months. It indicates that additional months of being in hospital due to psychiatric problem for one month or less and for about two month, reduces the hazards of schizophrenic patients by 57% and 63.5%, respectively. It takes some more months until the occurrence of an event by schizophrenic patients due to these risk factors. And the positive sign for the covariate estimate of the risk factor three or more months indicate that an additional month of being in hospital due to psychiatric problem for three or more months increases the hazard rate of schizophrenic patients by 3.6%. It is again equivalent to saying that schizophrenic patients who had been in hospital due to psychiatric problems for three or more months have fewer months until the occurrence of an event (death) compared with the one who had not been in hospital due to psychiatric problem.

The other important covariate that had been assumed to have a significant impact on the hazards of schizophrenic patients is SUBSTANCE ABUSE. The study shows that compared with the schizophrenic patients with no substance abuse, the estimated hazard rate of schizophrenic patient with a little and some substance abuse are respectively 0.354 and 0.613 times greater on average. This is to say that keeping other covariates at the same level the risk factor due to little substance abuse and some substance use increases the hazards of schizophrenic patients by 64.3 and 38.6 percent respectively compared with no substance abuse at all.

The odds ratio for a little and some substance use are both less than one indicating that compared to no substance use schizophrenic patient (the reference category), taking a little and some substance use reduces the survival longevity. Schizophrenic patients with no substance use took more months to experience an event. Based on the 95% confidence intervals of the hazard ratio, one can also interpret the results as using a little substance reduces the hazard rate from 76.3% to 47.3% and from 62.3% to 0.2% for users of some substance.

Table 11: Partial Likelihood Estimates for Fitted Proportional Hazards Models

(Stratification by Duration of illness)

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Marital			15.235	4	.004			
Marital(1)	-.627	.406	2.378	1	.123	.534	.241	1.185
Marital(2)	-.705	.414	2.899	1	.089	.494	.220	1.112
Marital(3)	-.183	.458	.159	1	.690	.833	.340	2.042
Marital(4)	-.015	.413	.001	1	.971	.985	.439	2.212
Substause			30.746	2	.000			
Substause(1)	-1.069	.201	28.285	1	.000	.343	.232	.509
Substause(2)	-.525	.248	4.495	1	.034	.592	.364	.961

Table 11 shows the partial likelihood estimate for the parameters of the proportional hazards model. In Table 11 the stratification is made with respect to the time-dependent covariates duration of illness. The interpretation for the resulting estimate can be made in similar way as we made in Table 9 and 10.

4.6 Test for the functional form of the model

Test for linearity in the relationship between the log of the hazards function and the covariates are one of the most common tests in the functional form for the Cox proportional hazards model. In most real situation an incorrectly specified functional form (non log-linearity) of the model covariate is a potential problem in the Cox regression model, especially for continuous variables.

LeBlanc and Crowley (1999) demonstrate that the average model error for a Cox model with incorrect covariate functional form is three times higher than a model that accounts for the non-linearity. Normally, it is assumed that covariates enter into the Cox partial log-likelihood in a log linear fashion, but continuous covariates can affect the hazard through more complicated nonlinear functional forms. Thus log-linearity in the functional form of the covariate is more likely to occur.

Failure to address the correct functional form of covariates in the model has serious consequences. If the hypothesized functional form is not same as the functional form we are using, the results obtained can change the interpretation of the effect and can radically alter the conclusions made from the model estimate. In that fitting the incorrect functional form is a form of misspecification and leads to some statistical errors: biased and decrease in the power of statistical significance. In general, the effects of incorrect functional form are similar to that of the failure of the proportionality of the proportional hazards model. Thus in order that the interpretation made from the model output is valid, the fit of the data with the model functional form has to be tested.

As it was described in Chapter Three the martingale residuals may be plotted against covariates to detect for the correctness of the functional form. The SPSS option of Cox-regression model allows us a model diagnostic with hazards function, partial residual and DfBeta(s). On SPSS the martingale residuals have to be recomputed as the difference of status variable and the cumulative hazards function. Then we plot the martingale residual against each covariates.

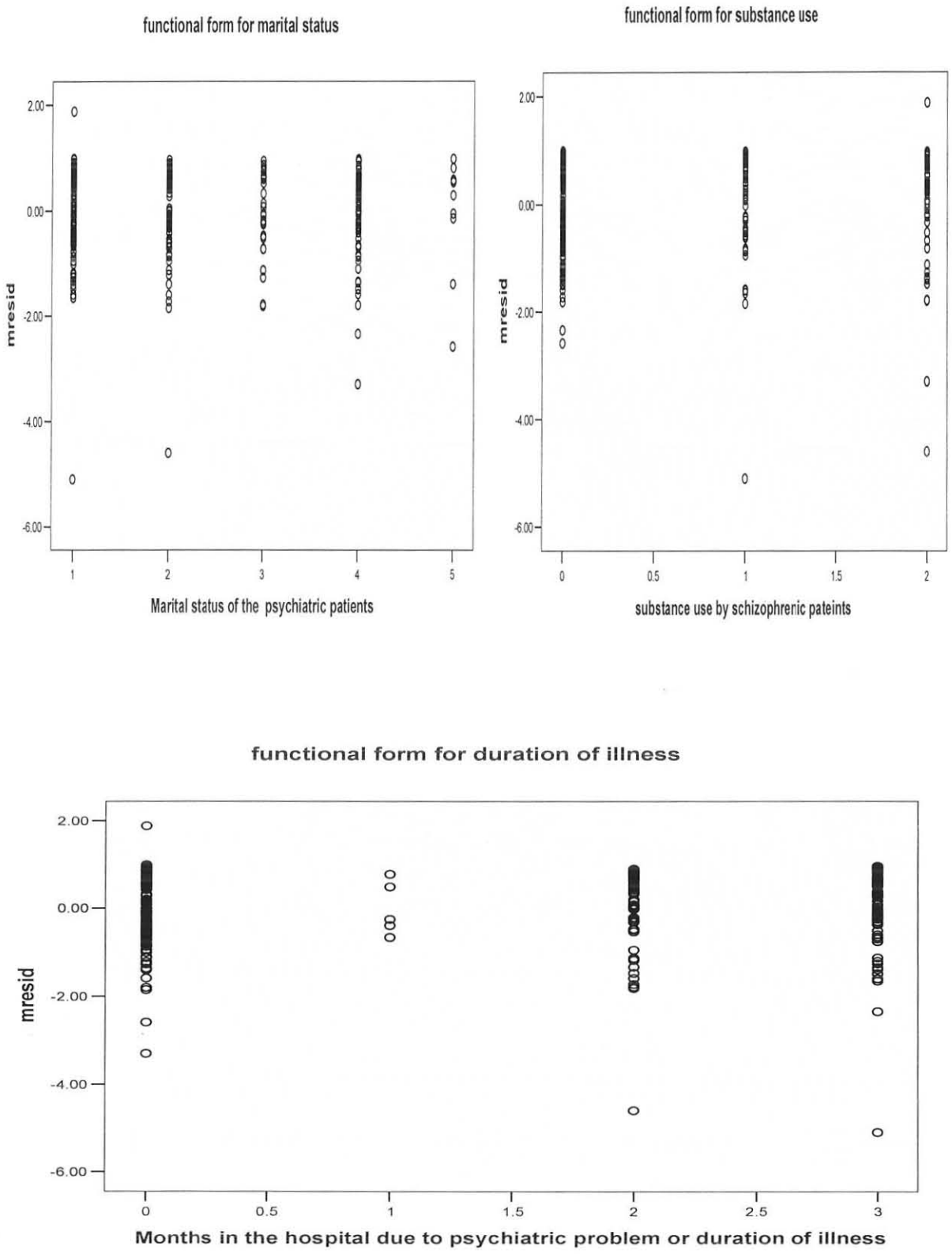


Fig 4: Figures for the assessment of functional form of the covariate in the model.

The panels of displays in Figure 4 show the plot of the martingale residual versus covariates. For each of the covariates, that is, marital status, substance use and the duration of illness, the plots reveal no evidence of any nonlinearity in the effect. The panels do not show a linear trend. In fact this is just for the sake of completeness since the problem of functional form is suspected about continuous covariates but not about categorical covariates like the ones we had considered. Irrespective of types of covariates (categorical or continuous) the panels show that the log linearity assumption of the covariate in the proportional hazards model is correct.

Chapter Five

Conclusions and Recommendations

5.1 Conclusions

The paper analyzes the effect of marital status, gender, substance use, duration of illness, physical illness, Year of education, age at the onset of illness and the current age of schizophrenic patient on their numbers of months of survival. The analysis was done based on a proportional hazards model.

The study suggests that marital status, duration of illness and substance use have statistically significant effects on the survival longevity of schizophrenic patients. On the other hand, physical illnesses, gender, age-at-onset of the illness and the current age of schizophrenic patient have no impact on the survival longevity of schizophrenic patients.

The study also assesses if the proportional hazards assumption for those statistically significant covariates is met. The covariates show the violation of the assumption both collectively and

individually. Thus we conclude that these covariates are time-dependent covariates whose effect will change from time to time over the study period.

In the assessment of the proportionality of the proportional hazards model assumption, the SAS program was used to check for the overall significance of the assumption and SPSS for individual tests of the assumption. The graphical test supported by statistical test was applied to check the validity of assumption. Up on rejection of the assumption an attempt was made to see the interaction effect between time and the time-dependent covariates. We finally use this model in analyzing the impact of the covariates on the survival longevity of schizophrenic patients.

The other important assumption about the proportional hazards model is the log linear assumption in the covariates. Even though this assumption is more likely to hold for continuous covariates, an attempt was made to assess the statistical significance of covariates in the model so as to confirm that the conclusion made is valid.

With regards to their marital status the married schizophrenic patients do have a better survival longevity than the never-married, separated, and divorced. For duration of illness due to psychiatric problems, schizophrenic patients who were never in a hospital due to psychiatric problem have a better survival longevity than those who have been admitted to a hospital for a month or less, for two months, and for three or more months. In the case of substance use, the study points out that compared to schizophrenic patients who never used a substance the probability of death of a schizophrenic patient with a little and some substance use is higher.

5.2 Recommendations

Based on the result of the study we forward some recommendation. The study shows that substance use has a statistically significant impact on the survival longevity of schizophrenic patients. We recommend that in order to reduce or stop substance use by schizophrenic patients the intervention of clinicians, governmental and non governmental bodies is necessary.

Alternatives to substance use for coping with stressful situations can be suggested. The reasons for substance misuse including relationship to psychiatric symptoms, antipsychotic treatment and feelings of social isolation must be explored, and then appropriate remedies can be given to patients by adopting concrete problem-solving approach. It is advisable to set tasks that are simple and readily achievable and encourage participating in activities with non-substance using peer group. The other important survival time determining factors are marital status and duration of illness. The study shows that being never-married reduce the hazards rates than in the case of separated, divorced and widowed. We strongly recommend that some orientation that will discourage separation or divorce for schizophrenic patients. As the number of months in hospital due to psychiatric problem increases the hazards rate increases as well. The physicians can advice the psychiatric patient to stay in the community.

REFERENCES

- A 2001 Institute of Medicine report on mental illness in developing country. http:// www. nap.edu/catalog/10111.html.
- Alan, S. B. and Carlo, C. D. (1999). Treating substance Abuse Among patients with schizophrenia. American Psychiatric Association, 50, 75-80.
- Allison, P. D. and Liker, J.K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. Psychological Bulletin, 91, 393-403.
- Allison, P.D. (1984). Event history analysis: Regression for longitudinal event data (Sage University paper series on quantitative applications in the social sciences, Number 07-046), Beverly Hills, CA: Sage.
- Allison, P.D. (1995). Survival Analysis using the SAS system: A practical Guide, Cary NC: SAS Institute.
- American National Centers for Health (1988). Current estimate from a National Health interview survey, United states 1987 vital and Health Statistics series, 10, No 166 Washington, DC. U. S Governmental Printing office.
- Andersen, P. K. (1982). Testing Goodness of Fit of Cox's Regression and life model. Biometrics, 38, 67-77.
- Arjas, E. (1988). A Graphical Method for Assessing Goodness of Fit in Cox's Proportional hazard Model. Journal of the American statistical Association 83, 204-212.
- Bennett, P. (1975). Techniques in industrial therapy, ergo therapy and reccreative methods. In Psychatric der Gegenwart (2^{nd} ed.), 3, 743-778.
- Bimerew, M. S., Soon, F.C .T. and Kortenbout, W.P. (2007). Substance abuse and the risk of readmission of people with schizophrenia at Amanuel psychiatric Hospital, Ethiopia: Curationis. 30(2):74-81.
- Bloessfeld, H.P., Hamerle, A. and Mayer, K. U. (1989). Event history analysis: Statistical theory and application in the social sciences. Hillsdale, NJ: Erlbaum.
- Box-Steffensmeier, J.M and Zorn, J. W. (2001). Duration Models and Proportional hazards in political science. American Journal of political science.45, 951-967.
- Breslow, N. (1974). Covariance Analysis of Censored survival data, Biometrics 30, 89-99.

- Brown, S., Inskip, H.and Barrarclough, B. (2000). Causes of the excess mortality of schizophrenia. British Journal of psychiatry; 177, 212-217.
- Brundland, H. (2001). Mental Health Care in the Developing World. Psychiatric Times, 19, Issue 1.
- Bund, G. R., Drake, R. and Mueser, K. T. (1997). An update on supported employment for people with severe mental illness. Psychiatric Services, 48, 335-346.
- Chapman, L. J., Day, D. and Burnstein, A. (1961). The process reactive distinction and prognosis in schizophrenia. Journal of Nervous and Mental Disease, 133, 383-394.
- Coetho, R.J. (1984). Self-efficiency and cessation of smoking. Psychological Reports, 54, 309-310.
- Collette, D. (1994). Modeling survival data in medical research. London: Chapman and Hall.
- Condiotte, M.M. and Lichtenstein, E. (1981). Self-efficiency and relapse in smoking cessation programs. Journal of Consulting and Clinical Psychology, 49, 648-658.
- Conwell, Y. and Caine, E.D. (2007). Mortality and age. The Royal College of Psychiatry, 190: 237-242.
- Cox, D.R. (1972). Regression models and life Tables (with Discussion). Journal of the Royal statistical society. Series B (Methodological), 34, No.2, 187-220.
- Cox, D.R. (1975). Partial Likelihood Biometrika, 62, 269-276.
- Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data. London: Chapman and Hall.
- Crowther, R. E., Morshall, M., Bond, G. R. and Huxley, P. (2001). Helping people with Severe Mental illness to Obtain Work: British Medical Journal; 322, 204-208.
- Derege Kebede, Atalay Alem, Teshome Shibre, Alemayehu Negash, Negussie Deyassa and Tefera Beyeru. (2004). The sociodemographic correlates of schizophrenia in Butajira, rural Ethiopia. Schizophrenia Research, 69 Issues 2-3, 133-141.
- Dixon, L., Hass, P.W., Weiden, P., and Frances, A. (1990). Acute effects of drug abuse in Schizophrenic patients: Clinical observations and Patients self-report. Schizophrenia Bulletin, 16, 69-79.
- Drake, R. E. and Wallach, M. A. (1989). Substance abuse among the chronically mental ill. Hospital and Community Psychiatry, 40, 1041-1046.

- Dress, K.A. (1986). The effect of gender identity on conversation. Social Psychology Quarterly, 49, 294-301.
- El-Guebaly, N. (1990). Substance abuse and mental disorders: the dual diagnosis concepts. Canadian Journal of Psychiatry, 35, 261-267.
- Everitt, B. (1987). Statistics in psychiatry. Statistical science, 2, 107-134.
- Farina, A., Garmezy, N. and Barry, H. (1963). Relationship of Marital status to incidence and Prognosis of Schizophrenia. Schizophrenia Research, 67, 624-630.
- Femalee, D. and Eder, D. (1983). Contextual effects in the classroom: The impact of ability groups on student attention. Sociology of Education, 56, 77-87.
- Femalee, D., Eder, D. and Tsui, W.Y. (1985). Peer influence on classroom attention. Social Psychology Quarterly, 48, 215-226.
- Fergusson, D.M., Horwood, L.J. and Dimond, M.E. (1985). A survival analysis of childhood family history. Journal of Marriage and the Family, 47, 287-295.
- Fergusson, D.M., Horwood, L.J. and Shannon, F.T (1984). A proportional hazards model of family background. Journal of Marriage and the Family, 46, 539-549.
- Fichman, M. (1988). Motivational consequences of absence and attendance: Proportional hazards estimation of a dynamic motivation model. Journal of Applied Psychology, 73, 119-134.
- Fleming, T. and Harrington, D. (1991). Counting Processes and survival Analysis. John Wiley and Sons, Inc., New York
- Foorman, S. and Lloyd, C. (1986). The relationship between social support and psychiatric symptomatology in medical students. Journal of Nervous and Mental Disease, 174, 229-239.
- Galonter, M. and Castaneda, R. (1988). Substance abuse among general psychiatric patients: place of presentation, diagnosis and treatment. American Journal of Drug and Alcohol Abuse, 14, 211-235.
- Gardner, W. and Griffin, W.A. (1989). Methods for the analysis of parallel streams of continuously reported social behaviors. Psychological Bulletin, 105, 446-455.
- Garfield, S. L. and Sundland, D. M. (1966). Prognostic scales in schizophrenia. Journal of Consulting Psychology, 30, 18-24.

- Garside, R. F. and Roth, M. (1978). Multivariate statistical methods and problems of classification in psychiatry. British Journal of Psychiatry, 133, 53-67.
- Geoff, D. and Everitt B. S. (2006). Statistical Analysis of medical data using SAS. Chapman and Hall, Now York
- Gove, W.R., Hughes, M. and Style, C. (1983). Does marriage have positive effects on the psychological well-being of an individual? Journal of Health and Social Behavior, 24, 122-131.
- Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnosis based on weighted residuals, Biometrika 81, 515-526.
- Greenhouse, J.B, Stangle, D. and Bromberg, J. (1989). An introduction to the survival analysis method for the analysis of clinical trial data. Journal of Consulting and Clinical Psychology, 57, 536-544.
- Gross, A. J. and Clark, V.A. (1975). Survival distribution. John Wiley and Sons Inc., New York.
- Hall, M., Adler, L., Leonard, S., and Freedman, R. (1979). Evidence in Postmortem brain tissue for decreased number of hippocampal nicotinic receptors in Schizophrenia. Biological Psychiatry, 38, 22-23.
- Harrison, G., Hopper, k., Graig, T., et al (2001). Recovery from Psychotic illness: a 15- and 25-year international follow-up study. British Journal of Psychiatry, 178, 506-517.
- Heckman, J. and Singer, B. (1985). Longitudinal analysis of labor market data. New York: Cambridge University press.
- Hembroff, L. A. and Myers, D. E. (1984). Status characteristics: Degrees of task relevance and decision processes. Social Psychology Quarterly, 47, 337-346.
- Jablensky, A., Sarturious, N. and Emberg, G. (1992). Schizophrenia: manifestations, incidence and course in different cultures: a World Health Organization 10-country study. Psychiatrical Medicine, Monograph supplement, 20, 1-97.
- Johnson, D. (1988). Panel analysis in family studies. Journal of the Marriage and family, 50, 949-955.
- Kalbfleish J.D and Prentice R. L (1980). The Statistical Analysis of Failure time data. John Wiley and Sons, Inc., Now York.

- Knudsen, P. and Wilmer, R. (1984). Cannabis and narcoleptic agents in schizophrenia. Acta Psychiatrica Scandinavica, 69, 162-174.
- Laska, E., Seigel, C. and Meilsner, M. (1985). Application of statistics to psychiatry. In Comprehensive Textbook of psychiatry. Department of Mental Health, Mental Retardation and Alcoholism Services, New York City, NY, USA.
- Lavori, P. W., Keller, M. B. and Klerman, G. L. (1984). Relapses in affective disorders: A reanalysis of the literature using life-table methods. Journal of Psychiatric Research, 18, 13-25.
- Lawless, J.F. (1982). Statistical Models and Methods for life time data. John Wiley and Sons, Inc., New York
- LeBlanc, M. and Crowley, J. (1999). Adaptive Regression Splines in the Cox Model. Biometrics 55:204-213.
- Limosin, F., Loze, J., Philippe, A., Casadebaing, F. and Rouillon, F. (2003). Ten-year prospective follow-up study of the mortality by suicide in schizophrenic patients. Schizophrenia Research, 94, Issue 1-3, 23-28.
- Litman, G.K., Eiser, J.R. and Taylor, C. (1979). Foundations of epidemiology (2^{nd} ed.). John Wiley, Sons, Inc., New York.
- Manning, C. and White, P. (1995). Attitudes of employers to the mental ill. Psychiatric Bulletin, 19, 541-543.
- Marks, J. (1963). Predicting outcomes in schizophrenia. Journal of Abnormal Psychology and Social Psychology, 68, 117-127.
- Mason, A. S., Tarpy, E. K., Sherman, S. J. and Haefrner, D. P. (1960). Discharges from mental hospital in relation to social class and other Variables. Archives of General Psychiatry, 2, 1-6.
- Mazberg, B. (1936). Marital status in relation to the prevalence of mental illness. Psychiatric Quarterly, 10, 245-261.
- McCullough, P. and Nelder, J. A. (1983). Generalized Linear Models: Chapman and Hall, New York.
- Miller, F., Busch, F. and Tanenbaum, J. (1989). Drug abuse in schizophrenic and bipolar disorders. American journal of Drug and Alcohol Abuse, 15, 291-295.
- Miller, R.G. (1981). Survival analysis. John Wiley and Sons, New York.

- Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). Introduction to the theory of statistics (3rd ed.). McGraw-Hill: New York.
- Moreau, T., O'Quigley, J. and Mesbsh, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. Applied statistics 34, 212-218.
- Moreav, T., O'Quigley, J. and Mesbah, M. (1986). On D. Schoenfeld's approach for testing the proportional hazards assumption. Biometrika 73, 513-515.
- Morgan, S. P., Lye, D. N. and Condran, G. A. (1988). Sons, daughters, and the risk of marital disruptions. American Journal of Sociology, 94, 110-129.
- Morita, J. G., Lee, T. W. and Mowday, R. T. (1989). Introducing survival analysis to organizational researchers: A selected application to turnover research. Journal of Applied Psychology, 74, 280-292.
- Mortensen, P. and Juel, K. (1993). Mortality and causes of death in first admitted schizophrenic patients. British Journal of Psychiatry, 163, 183-189.
- Muller, M. (2001). Goodness-of-fit Criteria for Survival Data. Institute for Medical Statistics and Epidemiology, IMSE. Technical University Munich, Munchen, Germany.
- Munk-Jorgensen, P. (1987). First admission rates and marital status of schizophrenics. Acta Psychiatrica Scandinavica, 76, 210-216.
- Nehrings, J., Hill, R. and Poole, L. (1993). Work, Empowerment and Community: Opportunities for People with long-term Mental Health Problems. London: RDP Publications.
- Nurris, V. (1956). A statistical study of the influence of marriage has on the hospital care of the mental sick. Journal of Mental Science, 102, 467-486.
- Nuttal, P. and Solomon, L. (1970). Prognosis in schizophrenia: The role of premorbid, social class, and demographic factors. Behavioral Science, 15, 255-264.
- Odegaard, O. (1946). Marriage and mental disease: A study in social psychopathology. Journal of Mental Science, 92, 35-39.
- Patel, V., Araya, R., Chatterjee, S., Chisholm, D., Cohen, A., De Silva, M., Hasman, C., McGuire, H., Rojos, G. and Van Ommeren, M. (2006). Treatment and prevention of mental disorders in low-income and middle income countries. The Lancet 370: 991-1005.

- Perkins, R. and Rinaldi, M. (2002). Unemployment rates among patients with long-term mental health problems. Psychiatric Bulletin, 26, 295-298.
- Philips, L. (1953). Case History Data and Prognosis in Schizophrenia. Journal of Nervous and Mental Disorders, 117, 515-525.
- Pocock, S. J. (1980). The role of statistics in medical research. British Journal of Psychiatry, 137, 188-189.
- Pokurny, A. and Overall, J. (1970). Relationships of psychopathology to age, sex, ethnicity, education and marital status in state hospital patients. Journal of psychiatric Research, 7, 143-152.
- Ponnuduria, R., Jayakar, J. and Sekaran, B. S. (2006). Assessment of mortality and marital ratio of schizophrenic patients a period of 30 years. Indian Journal of psychiatry, 48: 84-87.
- Pulver AE, Brown Ch, Wolynic J, Tam D, Adler L, Carpenter T and Childs B.(1990). Schizophrenia: age at onset, gender and Familial risk. Acta psychiatric scandinavica: Volume 82 Issue 5 page 344-351.
- Regier, D. A., Farmer, M. E. and Rae, D. S. (1990). Co morbidity of mental disorders with alcohol and other drugs of abuse: Results from the Epidemiological Catchments Area (ECA) study. Journal of the American Medical Association, 264, 2511-2518.
- Regier, D.A., Farmer, M.F., et al (1990). Co morbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) study. Journal of the American Medical Association, 264, 2511-2518.
- Rinaldi, M. and Hill, R. (2000). Insufficient Concern. London: Merton Mind.
- Robertson, N. C. (1974). The relationship between marital status and the risk of psychiatric referral. British Journal of psychiatry, 124, 191-202.
- Rowland, L. A. and Perkins, R. E. (1988). 'You can't eat, drink or make love eight hours a day'. The value of work in psychiatry Health Trends, 20, 75-79.
- Schizophrenia information. http:// www.Zyprexa.co.nz/schizophrenia/schizophrenia.html.
- Schoenfeld, D. (1982). Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 7, 145-153.

- Shapiro, D. R., Quitkin, F. M. and Fleiss, J. L. (1989). Response to maintenance therapy in bipolar illness: Effects of index episode. Archives of General Psychiatry, 46, 401-405.
- Shepherd, G. (1984). Institutional Care and Rehabilitation. London: Longman.
- Stevens, V. J. and Hollis, J. F. (1989). Preventing smoking relapse using an individually tailored skills training technique. Journal of Consulting and Clinical Psychology, 57, 420-424.
- Suvisaari, J., Haukka, J., Tanskanen, A. and Lonngvist, (1998). Journal Association of Familial leading with Early Onset of Schizophrenia, 173, 494-500.
- Taber, M.A. and Proch, K. (1987). Placement stability for adolescents in foster care: Finding from a program experiment. Child Welfare, 44,433-445.
- Teachman, J. D. (1982). Methodological issue in the analysis of family formulation and dissolution. Journal of Marriage and the Family, 44, 1037-1057.
- Therneau, T.R., Grambsch, P.M., and Fleming, T.R. (1990). Martingale based Residuals for Survival Models. Biometrika, 77, 147-160.
- Thompson L.A., Chhikera R.J. and Conkin J. (2003). Cox proportional Hazards Methods for Modeling the Time to Onset of Decompression Sickness in Hypobaric Environments. NASA Technical Publication 2003-210791, Houston Johnson Space Center.
- Tuma, N. B. and Hannan, M. T. (1984). Social dynamics: Models and Methods. San Diego, CA: Academic Press.
- Tuma, N. B., Hannan, M. T. and Groeneveld, L.P. (1977). Dynamic analysis of event histories. American Journal of Sociology, 84, 820-854.
- Turner, R. J., Dopkeen, L. S. and LaBreche, G. P. (1970). Marital status and schizophrenia: A study of incidence and outcome. Journal of Abnormal Psychology and Social Psychology, 76, 110-116.
- Walker, E., Bettes, B. A., Kain, E. L. and Harvey, P. (1985). Relationship of gender and marital status with symptomatology in psychotic patients. Journal of Abnormal and Psychology, 94, 42-50.
- Warner, R. (1994). Recovery from schizophrenia: Psychiatry and political Economy (2nd ed.). Oxford: Oxford University Press.
- World Health Organization (1979). Schizophrenic: an International Follow-up Study. Geneva, Switzerland: Author.

- World Health Organization (1998). Schedule for Clinical Assessment in Neuropsychiatry, Version 2-1. WHO.
- Yamoguchi, K. and Kandel, D. (1985). Drug and Alcohol Dependence, 50, 109-121.
- Yamoguchi, K. and Kandel, D. (1987). Drug use and other Determinants of premarital Pregnancy and its outcome: A Dynamic Analysis of Completing Life Events. Journal of Marriage and Family, 49, 257-270.
- Zatz, M. S. (1985). Los Cholos: Legal processing of Chicago gang members. Social Problems, 33, 13-30.

Appendix

Case Processing Summary

		N	Percent
Cases available in analysis	Event(a)	190	58.3%
	Censored	133	40.8%
	Total	323	99.1%
Cases dropped	Cases with missing values	3	.9%
	Cases with negative time	0	.0%
	Censored cases before the earliest event in a stratum	0	.0%
	Total	3	.9%
Total		326	100.0%

a Dependent Variable: survival time

Categorical Variable Codings (b, c, d, e, f)

		Frequency	(1)	(2)	(3)	(4)
physillnes(a)	0=not tall	296	1	0	0	
	1=a little	7	0	1	0	
	2=some	19	0	0	1	
	3=a lot	1	0	0	0	
DurIlliness(a)	0= none of the Period	185	1	0	0	
	1=1month	5	0	1	0	
	2=2month	52	0	0	1	
	3=3month	81	0	0	0	
Marital(a)	1=MARRIED	122	1	0	0	0
	2=NMARRIED	80	0	1	0	0
	3=SEPARATED	32	0	0	1	0
	4=DIVORCED	79	0	0	0	1
	5=WIDOWED	10	0	0	0	0
Gender(a)	1=MALE	210	1			
	2=FEMALE	113	0			
Substause(a)	0=not at all	221	1	0		
	1=a little	50	0	1		
	2=some	52	0	0		

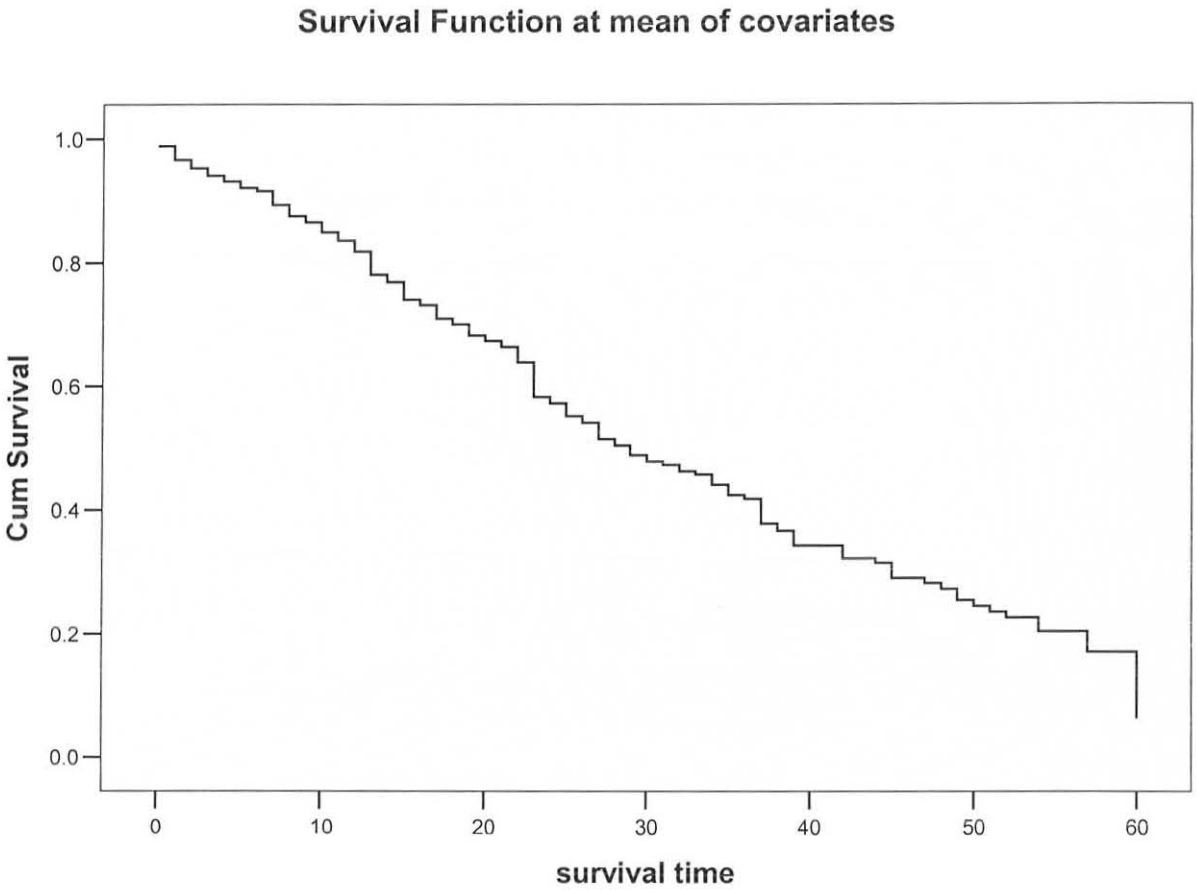
Survival Table

Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
0	.000	.977	.009	.023
1	.000	.966	.011	.034
2	.000	.953	.015	.048
3	.000	.941	.017	.061
4	.000	.931	.020	.071
5	.000	.921	.022	.082
6	.000	.916	.023	.088
7	.000	.894	.028	.112
8	.000	.875	.032	.133
9	.000	.866	.034	.144
10	.000	.849	.037	.163
11	.000	.836	.040	.179
12	.000	.819	.043	.200
13	.000	.781	.050	.248
14	.000	.769	.052	.263
15	.000	.740	.057	.301
16	.000	.732	.058	.312
17	.000	.710	.062	.343
18	.001	.701	.063	.356
19	.001	.683	.066	.382
20	.001	.673	.067	.396
21	.001	.664	.068	.410
22	.001	.639	.072	.448
23	.001	.583	.078	.540
24	.001	.572	.079	.558
25	.001	.552	.081	.595
26	.001	.541	.082	.615

27	.001	.515	.084	.664
28	.001	.504	.085	.685
29	.001	.489	.086	.716
30	.001	.478	.087	.738
31	.001	.473	.087	.748
32	.001	.463	.088	.771
33	.001	.457	.088	.782
34	.001	.441	.089	.819
35	.001	.424	.089	.858
36	.001	.418	.089	.872
37	.001	.378	.090	.973
38	.001	.367	.090	1.004
39	.002	.343	.090	1.071
42	.002	.322	.089	1.134
44	.002	.315	.089	1.156
45	.002	.290	.088	1.238
47	.002	.282	.088	1.267
48	.002	.273	.087	1.300
49	.002	.254	.086	1.370
50	.002	.245	.085	1.406
51	.002	.236	.084	1.446
52	.002	.226	.084	1.487
54	.002	.204	.081	1.591
57	.003	.171	.077	1.767
60	.004	.063	.043	2.767

Covariate Means

Variable name	Mean
Level of education in years	15.012
age1	23.632
physillnes(1)	.916
physillnes(2)	.022
physillnes(3)	.059
Durllliness(1)	.573
Durllliness(2)	.015
Durllliness(3)	.161
Marital(1)	.378
Marital(2)	.248
Marital(3)	.099
Marital(4)	.245
Gender	.650
age2	30.604
Substause(1)	.684
Substause(2)	.155

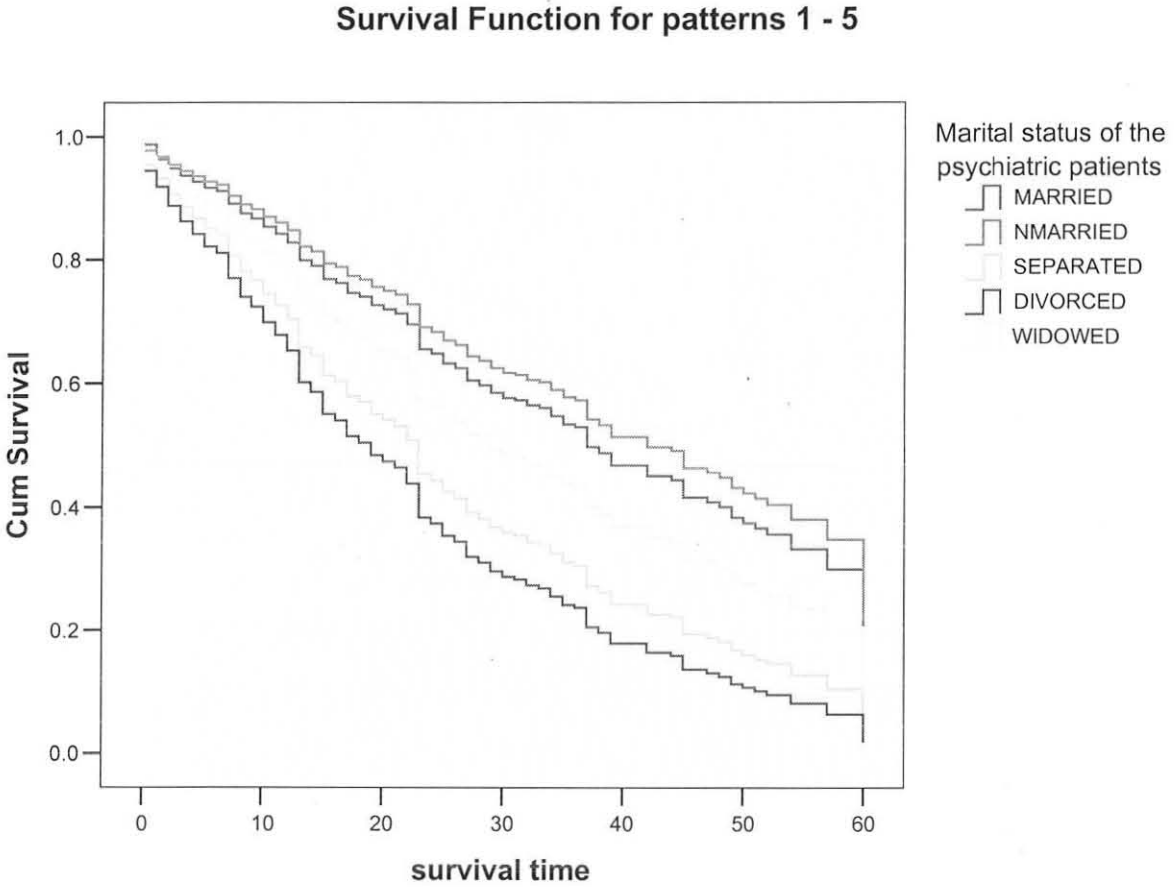


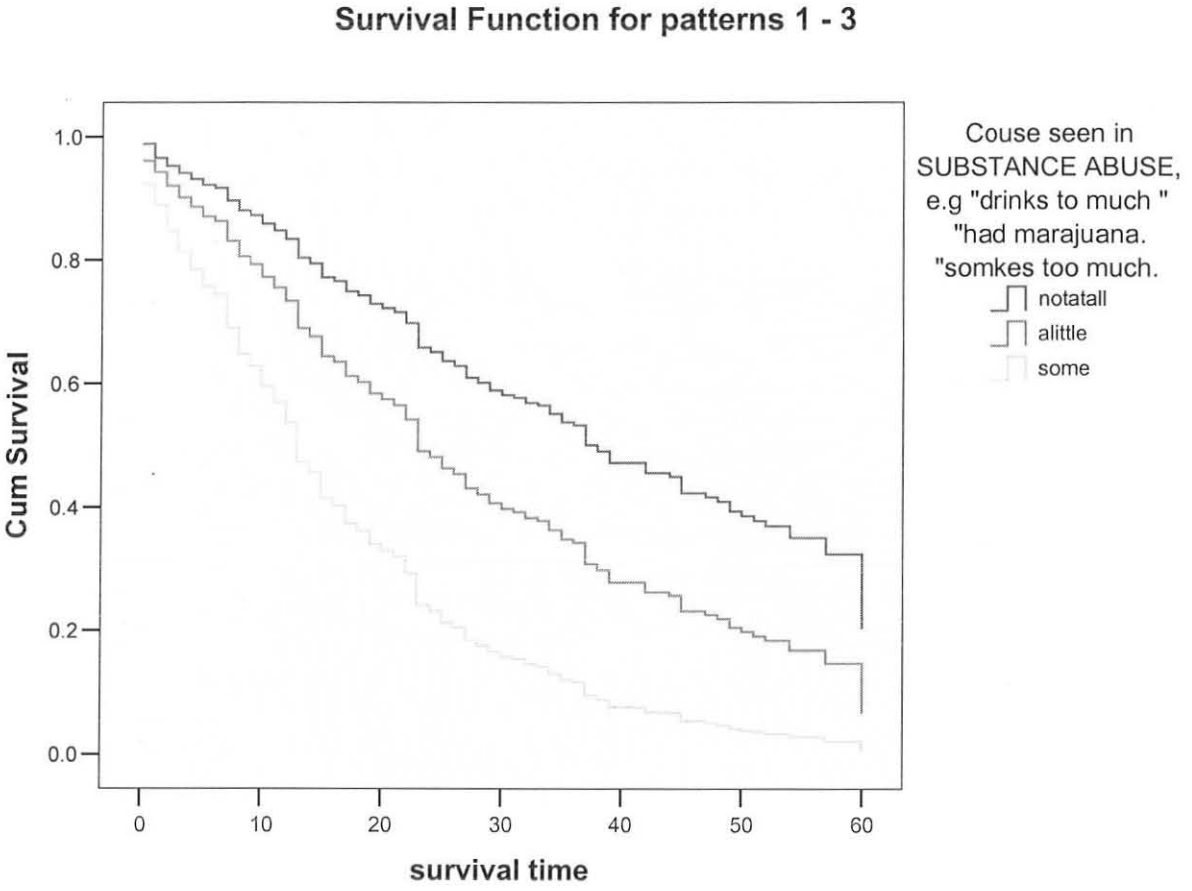
SAS code for testing proportional hazards assumption

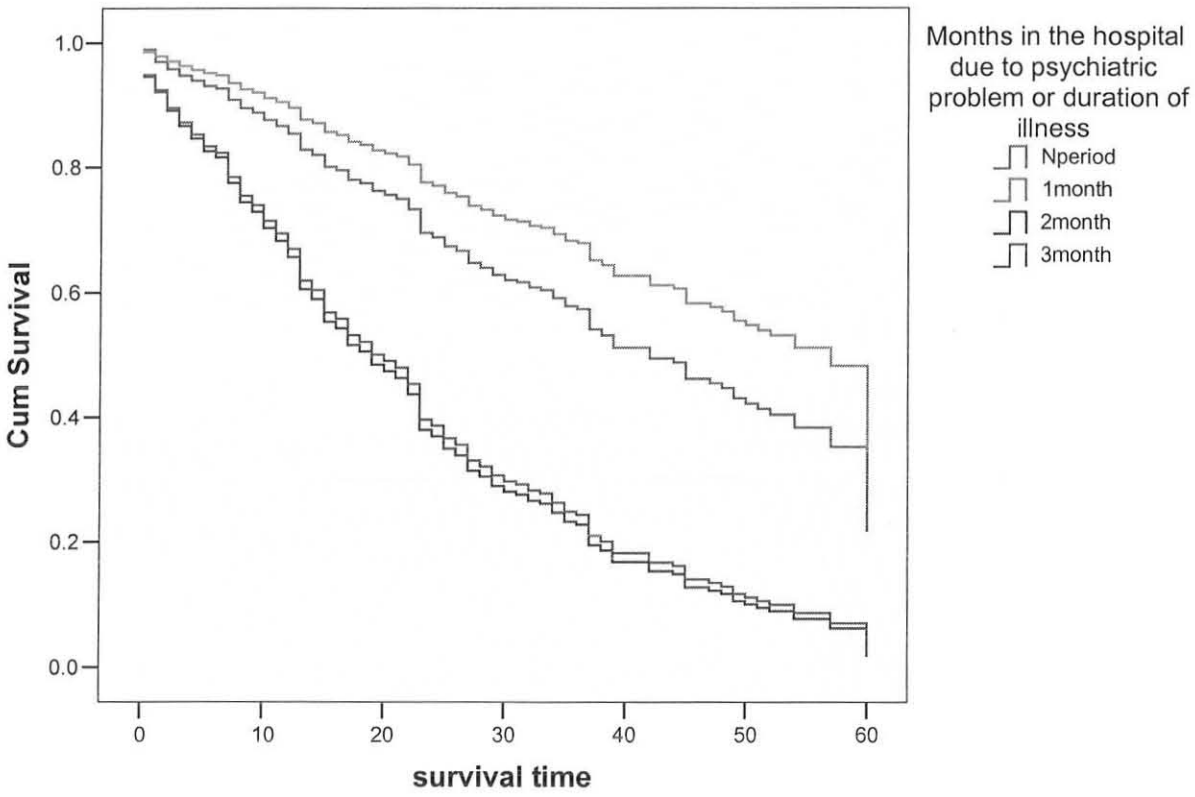
```
data schizo;
Set work schizo;

Proc phreg data=schizo;
model time*status (1) =dur mar subuse durt mart subuset;
durt=dur*LOG (time);
mart=mar*LOG(time);
subuset=subuse*LOG(time);
proportionality_tet: test durt, mart, subuset;
run;
```

Survival curve for each of the statistically significant covariate




Survival Function for patterns 1 - 3
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
survival time
Couse seen in SUBSTANCE ABUSE, e.g "drinks to much " "had marajuana. "somkes too much.
notatall
alittle
some


Survival Function for patterns 1 - 4
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
survival time
Months in the hospital due to psychiatric problem or duration of illness
Nperiod
1month
2month
3month