STRUCTURAL ANALYSIS AND MODELLING OF ALCOHOLISM AS A NON-COMMUNICABLE DISEASE

BY

GEORGE MWANGI MURIUKI

A THESIS SUBMITTED TO THE SCHOOL OF SCIENCES AND AEROSPACE STUDIES, DEPARTMENT OF MATHEMATICS, PHYSICS & COMPUTING IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOSTATISTICS

MOI UNIVERSITY

2022

DECLARATION

Declaration by Candidate

This thesis is my original work and has not been presented for a degree in any other University. No part of this thesis may be reproduced without the prior written permission of the author and/or Moi University.

Signature:_____ Date_____

George Mwangi Muriuki

DPS / PHD/ 008 / 016.

Declaration by the Supervisors

This thesis has been submitted with our approval as University supervisors:

Signature:_____ Date_____

PROF. JOHN MUINDI MUTISO

Department of Mathematics Physics and Computing

School of Sciences and Aerospace Studies

Moi University

Signature:_____ Date_____

DR. MATHEW. K .KOSGEI

Department of Mathematics, Physics and Computing

School of Sciences and Aerospace Studies

Moi University

DEDICATION

To my son (Rafiq Amos Muriuki Mwangi) and daughter (Patience Wairimu)

No eye has seen, no ear has heard what God has prepared for you.

ACKNOWLEDGEMENTS

First I thank God for granting me a truly life-changing experience while undertaking this PhD. Secondly, my most profound appreciation goes to my esteemed supervisors Dr Mathew Kosgei and Professor John Mutiso for their time, effort, and understanding in helping me succeed in my studies. Their vast wisdom and wealth of experience have inspired me throughout my studies. In addition, I am greatly indebted to Dr Mathew for his treasured technical assistance throughout my research.

I'd like also to thank my PhD coursework lecturers, Prof Koske and Prof Kinyanjui for their advice, and my classmates; Andika, Beatrice, Beryl, Celestine, Hellen, Jemimah, Kiplimo, "Kijana", Muthui and Toromo for encouragement and the time we spent together.

I greatly appreciate Mr Cheburet, Ms Grace and Mr Warutere from the Ministry of Health, Morris from NACADA, the R-studio bloggers, bar operators, owners and all those rehabs, who generously and sincerely provided me with information that I needed.

Words cannot express my gratitude to Prof L. Odongo and Mr Riungi,(former principal Gituamba Sec Sch) .

I gratefully thank National Research Fund for the funding opportunity to undertake my research. Many thanks goes to Dr. Silver and entire Statistics department, Moi University for the support.

I am indebted to Richard Songoro whose remark during the master's program changed my academic trajectory. My deep appreciation go to Prudence for immense encouragement. In the end, I am grateful to my parents, siblings, friends and acquaintances who remembered me in their prayers for the ultimate success. You all made my academic journey memorable.

ABSTRACT

Alcohol use has been part of many cultures globally for many years. However, alcoholism or alcohol use disorders was recognized in Kenya in 2014 as a disease as classified by World Health Organization. Alcoholism is one of the highest causes of global disease burden resulting from liver cirrhosis, road traffic accidents and several types of cancers. In Kenya alcoholism is a persistent health problem with harmful alcohol consumption, especially among young people (18-35 years), being on the rise in spite of stringent laws governing alcohol use. Existing models consider transmission of alcoholism through social interaction in addition prevalence and incidences of alcohol use and alcoholism in Kenya have mainly been determined using surveys. Models that address the progression stages of alcoholism, from susceptible through social drinking to alcoholism are lacking. This study sought to analyse structural relationship between risk factors and alcoholism and model alcoholism as a non-communicable disease. The specific objectives were to: Analyse structural relationship between risk factors and alcoholism; model incubation period of alcoholism; evaluate parametric and non-parametric hazard of alcoholism and predict incidences of alcoholism. **Method:** Secondary data was sourced from the Ministry of Health. Structural Equation Modelling (SEM) was used to show structural relationship between alcoholism and the latent variables; to model incubation period of alcoholism Birnbaum-Saunders (B-S) distribution hinged on biological process of cumulative cell damage caused by excessive alcohol consumption was used; hazard of becoming alcoholic was determined non-parametrically using discrete-time hazard model and incidences of alcoholism were found using logistic regression and back-projection. **Results:** There was a significant relationship between alcoholism and the risk factors (gender, peer influence, age at onset, social-cultural, economic status, environmental settings, drinking habits and pattern, family and family attention and personality) (RMSEA =0.06; CFI =0.80; SRMR=0.06); B-S distribution (α =0.77 [CI: 0.68, 0.85] and β = 6.13 [CI: 5.44, 6.83], R^2 =0.94, was appropriate for modelling incubation period of alcoholism; the probability of becoming alcoholic increased from 0.31% when drinking once per week to 57% when drinking seven sessions a week, in addition the hazard of becoming alcoholic was higher for females than for males; The model was used to predict incidence of alcoholism between 2014 and September 2019. The predicted number of alcoholics (6632) do not differ significantly from the reported cases alcoholics (6631). In conclusion the analyses of the relationship between risk factors and alcoholism showed that risk of becoming alcoholic was affected differently by different risk factors with gender having the largest impact. Biophysical process of fatigue failure caused by cumulative cell damage yielded a model of alcoholism as a non-communicable disease. **Recommendation:** The study recommends initiatives for sensitization on impacts of alcohol use and early diagnosis of alcoholism to help initiate prevention policies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

AAF	Alcohol-Attributable Fraction
ALD	Alcohol Liver Disease
AIC	Akaike Information Criterion
ASTM	American Society for Testing and Materials
AUD	Alcohol Use Disorder
AUDIT-C	Alcohol Use Disorders Identification Test consumption questions
B.A.C	Blood Alcohol Concentration
BMI	Body Mass Index
B-S	Birnbaum and Saunders distribution
CAGE	need to Cut down, Annoyed by criticism, Guilty after drinking, need for Eye-opener in the morning
CFA	Confirmatory Factor Analysis
CIH	Center for Integrated Healthcare
DSM –IV	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
ICD-10	International Classification of Diseases, 10th edition
NACADA	National Authority for the Campaign against Alcohol and Drug Abuse
NCD	Non-communicable Diseases
RMSEA	Root Mean Square Error of Approximation
S.E.S	Social Economic Status
S.I.R	Susceptible, Infectious and Removed states
SEM	Structural Equation Modelling
SRMR	Standard Root Mean Square Residual
TLI	Tucker Lewis Index
UNAIDS	Joint United Nations Programme on HIV/AIDS
W.H.O	World Health Organisation

DEFINITION OF TERMS

Alcohol: Alcohol is "a colourless volatile liquid obtained by fermenting sugars and starches with yeast and is used as a solvent, skin cleaner, hardener, and as an intoxicating drink", Youngson (1999).

Alcoholic: An alcoholic is "a person who habitually consumes alcoholic drinks to excess or who is addicted to alcohol to the detriment of health. A person suffering from alcoholism", Youngson (1999).

Alcoholism: Alcoholism is "(a) psychological dependence on alcohol with compulsive consumption of alcohol, (b) damage to the stomach lining; liver, nervous system, heart or voluntary muscles caused by prolonged exposure to high blood levels of alcohol". Also known as the ''alcohol dependence syndrome,' or "alcohol use disorder". Both ICD 10 and DSM-IV differentiate harmful alcohol use or abuse and dependence respectively**.** In this thesis, the terms alcoholism and Alcohol Use Disorder (AUD) have been used interchangeably as generic term referring to both (a) or (b) or both, Youngson (1999).

Incubation period. The incubation period is the time from pathogen exposure to the onset of symptoms, in relation to infectious diseases. Or time from exposure to development of symptomatic disease, in relation to non-communicable diseases.

Susceptible. Those people who do not consume alcohol and have never consumed alcohol but are at risk of taking it.

Social drinkers. These are people who drink moderately (one who takes 4 to 14 drinks per week and no more than 2 at once) and have no problems associated with alcohol use. They include casual drinkers and are also known as moderate drinkers.

Non-communicable diseases. These are health conditions or diseases which are "non-infectious". They are diseases that cannot be spread from one person to another. Also called non-transmissible diseases.

Asymptomatic. Having no symptoms of illness or disease i.e. symptomless.

Event. The term "event" in event history analysis represents a change or transition from one state or condition of interest to another, Box-Steffensmeier & Jones, (2004).

Onset. Beginning of regular intake of average amount of alcoholic beverages.

A **model** is a statistical statement about the relations among variables.

Fatigue. Fatigue is a process of progressive, localized, and structural change occurring in a material subject to conditions that produce fluctuating levels of stress and strain. Or fatigue is a cumulative damage process presented in materials subject to periodic and fluctuating stress levels during each duty cycle, ASTM (2013).

Communicable disease. Means an illness due to a specific agent or its product that arises through transmission of that agent or its product from an infected person, animal, anthropoid or inanimate reservoir to a susceptible host, either directly or indirectly, through an intermediate plant or animal host, vector, or the inanimate environment.

One unit of alcohol or a drink. One unit equals **10ml or 8g of pure alcohol**, which is around the amount of alcohol the average adult can process in an hour. This means that within an hour there should be, in theory, little or no alcohol left in the blood of an adult, although this will vary from person to person. It is equivalent to:

A single measure of spirits (ABV 37.5%); half a pint of average-strength (4%) lager; two-thirds of a 125ml glass of average-strength (12%) wine; half a 175ml glass of average-strength (12%) wine; a third of a 250ml glass of average-strength (12%) wine.

One session of alcohol consumption. This was defined as a non-interrupted drinking of alcohol separated to the next session by at time of at least four hours. **Note**

The half-life of alcohol is **four to five hours**. A half-life is how long it takes for your body to get rid of half of it. But you need about five half-lives to get rid of alcohol completely.

Chronic A chronic condition is a health condition or disease that is persistent or otherwise long-lasting in its effects or a disease that comes with time. The term chronic is often applied when the course of the disease lasts for more than three months. Wikipedia

Disorder an abnormal physical or mental condition ("Merriam-Webster", n.d) https://www.merriam-webster.com/dictionary/disorder

Disease. Literally the lack of ease, or illness /suffering, Panda (2009). **Or**

A disease is a particular abnormal condition that negatively affects the structure or function of all or part of an organism, and that is not immediately due to any external injury. Diseases are often known to be medical conditions that are associated with specific signs and symptoms. Wikipedia

Risk factor. This is a measurable characterization or variable of each subject that significantly predicts whether an individual will develop a disorder or disease. Donovan, (2004).

Risk period is the period when one is exposed to a certain risk of a particular event. e.g. only persons who are drinking can experience alcoholism. Thus they are at risk

The **hazard rate** is the frequency at which the event of interest occurs per unit of time, given that it hasn't yet happened

States. These are the categories of the `dependent' variable and one can only be in one state at any given time. For example either alcoholic or not

CHAPTER ONE: INTRODUCTION

1.0 Introduction

This chapter contains background of the study, problem statement, rationale of the study, study objectives, significance and limitations of the study.

1.1 Background Information

Globally, alcohol consumption is well entrenched in the social fabric of many adult populations, virtually constituting a behavioural norm. Alcohol is readily available and mostly cheap and its use in Kenya is legal under the Alcoholic drinks control Act. Sustained alcohol consumption is a brain-centred addictive behavioural disorder that transcends all boundaries of gender, race, age and economic strata leading to alcohol dependency in some people, (Bloom et al., 2011; Rehm et al., 2012; WHO, 2014).

Recent studies give overwhelming evidence that the increase in the risk of alcohol-related harm, mainly in a dose-dependent manner, results from a combination of frequency of drinking and amount drunk per incident and the volume of lifetime alcohol use (Rehm et al., 2010a; WHO Regional Office for Europe, 2009) confirming that higher alcohol consumption causes the greater risk.

In Kenya, which is a developing country with a relatively young population, alcoholism is on the rise in spite of stringent laws by both the central and county governments on sale and consumption of alcohol. Alcoholism pose a serious and persistent health problem with harmful alcohol consumption, especially by the persons aged between 15 and 35 years, (Baridi, 2014). In 2020, at height of the Coronavirus pandemic local dailies published several articles on the Kenyan drinking culture especially the fact that a large number of alcohol users were having a challenge in observing the Ministry of Health guidelines regarding alcohol use. In April 24 2020, The Nairobian carried a

heading “Kenya is a drinking Nation”. It was noted that Nairobi city has more bars than New York City, (Odongo, 2020). Normally, majority of Kenyans drink from bars, clubs, wines and spirits shops, and a few from home (Mahugu, n.d.) with addicted ones taking it shortly after waking up before engaging in any productive activity.

People get initiated into alcohol use for different reasons. They are (un)aware of the risk in the adventure. Many persons who take their first alcoholic drink at an early age (before 14 years) have a high chance of becoming alcoholic within 10 years of inception (Hingson et al. (2006). (DeWit et al., 2000 and Gómez et al., 2011) noted that onset of alcohol use between ages 11–14 greatly increases the risk of development of alcohol use disorders.

Alcohol use is one of the highest risk factors for disease and disability, after childhood underweight and unsafe sex. It contributes to traumatic outcomes that kill or disable people at a relatively young age, resulting in the loss of many years of life to death and disability and causing approximately 4.5% of the global burden of disease and injury, (Bloom et al., 2011; Rehm & Sempos, 1995; White& Gasperin, 2006; WHO, 2014; Wilson et al., 2004; Wilson et al. 2010).

Alcohol consumption is a causal factor in more than 60 major types of diseases, conditions and injuries such as; cardiovascular diseases, cancers, lung diseases, gastrointestinal conditions, mental and behavioural disorders, reproductive disorders , immunological disorders pre-natal harm, low birth weight, skeletal and muscular diseases, Anderson & Baumberg (2006). This results in approximately 2.5 million deaths globally each year. This number is predicted to rise especially in developing countries.

Alcohol abuse is a serious condition which results in severe negative consequences for addicts, their friends and families, and the community at large. Heavy drinking also significantly increases morbidity and mortality from infectious diseases. Alcoholism takes different forms with researchers suggesting that different subtypes of alcoholism may have different aetiologies, Cloninger et al. (1996); Zucker et al. (1996).

The term alcoholism was first coined in 1849 by Magnus Huss (Baridi, 2014). Historically alcoholism was considered a sin, a moral failing, a character defect, and only recently as a disease, it also is one of most misunderstood illnesses (Gordis, 1997). There are two views on alcoholism; one that considers alcoholism as an independent disease as defined comprehensively by (Youngson, 1999) and the other views it as a symptom of an underlying disorder. WHO (2016), contains ICD-10 guidelines on disease categorization. Alcoholism spans several non-communicable diseases categories such as F10-F19 which covers mental/psychological symptoms among others (containing conditions resulting 100% from alcohol use, i.e. AAF = 100%). These categories have been summarised in Appendix A.

Non-communicable diseases (NCDs) are a major health problem of this century. World over NCDs have imposed enormous individual and social burden that will even increase in the coming decades, Brinks and Landwehr (2014). Mortality and morbidity arising from non-communicable diseases also called silent killer are on the rise, (Bradshaw et al., 2011; WHO, 2002). Main cause of NCDs includes environmental, nutritional deficiencies, lifestyle choices and genetic inheritance. Alcohol use being a lifestyle choice.

Alcohol (the active ingredient in alcoholic drinks) is a colourless liquid with chemical formula CH_3CH_2OH. It is also known as ethanol or ethyl alcohol and is produced by

brewing, distillation and fermentation. Once taken, alcohol is rapidly absorbed along the entire gastrointestinal tract by passive diffusion without involving an active transport system and then distributed to all body organs especially those rich in oxygen such as brain, heart and lungs. Physical, psychological, and biological factors such as concentration, nature of alcoholic beverage, stomach contents and duration of drinking session interact with individual's gender, Body Mass Index (BMI), age, body water and personal habits to influence rate of alcohol absorption, (Dubowski, 1985; Alac, 2012). Body treats alcohol as poison triggering an elimination process immediately after absorption through metabolism mostly in the liver.

By exploring the issues surrounding alcoholism, readers will understand better, why, despite thousands of years of involvement with alcoholic beverages, alcohol related problems continue to plague human society (Pal & Ray, 2016).

1.1.1 The structural relationship between alcoholism and its risk factors

Issues of alcoholism are often complex and multidimensional in nature. The biological mechanisms underlying alcoholism are not clearly identifiable, however physiological factors such as heart rate, BMI and demographic variables such as sex, age and factors associated with lifestyle such as smoking history, dietary habits are known to affect time to alcoholism of a social drinker (Collett, 1994).

Since not everyone who consumes alcohol becomes alcoholic, then it means that there is a combination of factors that triggers the acquisition of this disorder. The factors that make some individuals become alcoholic and others not despite consuming large quantities of alcohol for a long period of time continue to intrigue researchers. These factors cannot be directly measured since alcoholism is a latent disorder in which only symptoms of the disorder are manifest either physically or psychologically. Forming

mathematical models to show structural relationships require a method that tests whether the conceptualized model fits the data and at the same time accounts for the latency property of alcoholism. One of the model that has these qualities is Structural Equation Modelling (SEM).

SEM is a statistical approach for testing hypothesis about the relationships among observed and latent variables (Hoyle, 1995). As a quantitative, second generation multivariate statistical analysis method, SEM combines the benefits of path analysis, factor analysis and multiple regression analysis.

SEM being a technique for investigating relationships between latent (unobserved) variables or constructs that are measured by manifest (observed) variables or indicators requires the separation of variables into two classes: observed and latent variables. Structural model assesses the direction and strength of the relationship between endogenous and exogenous variables. In model specification statements about a set of variables are formulated and then a diagram or pictorial representation of the model is transformed into equations which are then solved simultaneously to estimate parameters and test model fit.

The most important reason for rising use of this statistical technique is because direct and indirect relationships among causal variables can be assessed with a single model (Meydan & Şen, 2011). Its ability to take into account measurement errors and the relationships between errors in the observed variables minimize measurement errors (Civelek, 2018).

There are several research studies seeking relationship between alcohol use and its risk factors that have applied SEM.

1.1.2 Incubation period of alcoholism.

Alcoholism is a chronic disease with long duration and slow progression characterised by physical and or psychological dependence on alcohol (Nacoa, (n.d).). It has all characteristics of a non-communicable disease such as: complex aetiology, multiple risk factors, has non-contagious origin, is not communicable; causes premature morbidity, dysfunction and reduced quality of life (leads to functional impairment or disability) and usually develops and progresses over long periods (has a prolonged course of illness), WHO (2011). Once manifested there is usually a protracted period of impaired health. In addition it has long latency period and is often initially insidious

Alcohol is metabolised first to acetaldehyde (highly toxic substance), which is then broken down to less toxic compounds. If one drinks alcohol at a rate faster than that of alcohol oxidation then, the amount of alcohol in the blood or Blood Alcohol Concentration (BAC) keeps on rising up. High levels of alcohol in the blood cause cellular and finally organ damage, especially to the organs involved in the digestion process (such as liver) and the nervous system, (Gmel & Rehm, 2003; Rehm et al., 2003a and WHO, 2011).

Gradual cell damage resulting from regular/sustained alcohol consumption has been illustrated by Frazier et al. (2011); Gao & Bataller (2011); Ma & Brunt (2012) among others. This ultimately leads to weakened organs or functional impairment.

> Regular drinking changes the chemistry of the brain, especially depletion of the neurotransmitter serotonin. This leads to the cyclical process of drinking to relieve depression, becoming more depressed as levels of serotonin become more depleted, thus needing more alcohol to medicate the depression. (Cornah, 2006, p. 7).

Ultimately a social drinker may start showing signs of alcoholism. The duration between initiation to alcohol consumption and onset of symptoms is the incubation

period and varies between individuals. It was modelled using life distribution based on theoretical and empirical justifications.

1.1.3 The hazard of becoming alcoholic

To reduce risks associated with excessive/harmful alcohol use it is necessary to know which combination of explanatory variables affect the hazard of becoming alcoholic.

Hazard rate, is perhaps the most popular representative in modelling and analysis of lifetime data because of its intuitive interpretation as a risk by a unit to fail at age $\boldsymbol{x}$. According to Rinne (2014), the hazard rate is more informative about the underlying failure mechanism than other representatives of lifetime distribution. Hence it is the dominant method used in summarizing survival data. Besides, hazard is a 'single failure' system of the complete intensity function. Being a special case of intensity function for non-homogeneous Poisson process, hazard rate models occurrence of only the first event say onset of alcohol consumption (Rinne, 2014). Hazard rate measures propensity to fail. For a susceptible it measures propensity of starting to drink, while for a social drinker the propensity to become alcoholic. Hazard rate can be calculated using parametric and/or non-parametric methods.

Parametric methods assume that the time until an event occurs can be modelled by specific distribution for example Weibull distribution, Allison (1984). Nonparametric methods make very few assumptions if any about the distribution of an event time. Both methods can estimate the effects of covariates on hazard rates, Danelia (2011). Semi-parametric methods which combine both parametric and non-parametric properties (such as Cox-proportional hazards regression models) also exist. In this study, hazard rate was parametrically determined using Birnbaum-Saunders distribution and non-parametrically estimated using event history analysis.

Term survival analysis is used mostly in biomedical studies with event history analysis being used in social sciences where the phenomenon of interest is time-to-event or a duration, and/or the response is the occurrence of a discrete event in time.

Event history analysis is used to study the duration until the occurrence of the event of interest, where the duration is measured from the time at which an individual becomes exposed to the 'risk' of experiencing the event. Event history is also known as survival analysis, duration analysis or hazard modelling, failure-time models, reliability models. The name survival analysis or event history analysis emanates from studying how long subjects in a study survive under different circumstances, Allison (1984).

The implicit interest in event history analysis is survival and risk as something persists (in our case surviving as a social drinker). "Event history methods permit researchers to make claims not just about the factors that precipitate the risk but also, how differences are related to this risk" (Box-Steffensmeier & Jones, 2004, p4). At a minimum, event history data contain information on when the units begun the process under study and information on the timing of the event's occurrence (if an event is observed within the span of the observation plan). Hence, event history is a longitudinal record of the timing of the occurrence of one or more types of event. The hazard function is also used to estimate the survival function.

1.1.4 Projections of alcoholism incidences.

Back calculation has been a popular statistical approach in predicting the future of AIDS epidemic. It is used to reconstruct the historical infection rates which generated the observed pattern of diagnoses. Back-calculation method was proposed first by Brookmeyer and Gail, (1988). Later, Gail & Brookmeyer (1989) made projections on incidences of AIDS using this method. In back-calculation methodology, distribution

of incubation period is assumed to be exactly known (Ravanan & Venkatesan, 2008). Like the study of AIDS (which lack data on infection), study of alcoholism face the problem of availability of accurate data regarding onset of the alcohol drinking habit.

Back-calculation method has undergone several modifications and development over the years to make it applicable in many different situations. However, it has strengths as well as weaknesses. It has not been extensively used in modelling of NCDs.

Brinks & Landwehr (2014, p.62) noted that "Despite the importance of NCDs, mathematical models for the dynamics of NCDs are rarely examined. This is in contrast to infectious diseases with a variety of modelling approaches". This study sought to introduce the concept of incubation period of alcoholism in modelling alcoholism as a non-communicable disease. Effect of alcohol on cells during alcohol metabolism was applied to model time from onset of alcohol taking to diagnosis with alcoholism. Social interaction (the back-bone of deterministic model of alcoholism) was incorporated as one of the factors leading to alcoholism.

1.2 Problem Statement

In Kenya, alcohol has become alarmingly the most routinely used and misused substance among both adults and minors in the contemporary society and its socio-economic impact is devastating and on the rise, especially in the young persons (18-35 years), in spite of stringent laws governing sale and use of alcohol. The risk factors strongly linked to alcohol use and alcoholism have been studied disjointedly resulting in different structural models being proposed. In addition, existing mathematical models of alcoholism are deterministic and assume alcoholism is infectious in order to model the system where the alcoholics 'infect' those who are not drinking or 'susceptibles'.. Models that address the effects of alcohol on the body of the drinker in

order to show the progression to alcoholism from susceptible state through social drinking state are lacking. This study sought to undertake structural analysis of alcoholism and its risk factors and also model alcoholism as a non-communicable disease by applying the biophysical effect of alcohol on the body of the consumer which cause cumulative cell damage as alcohol use progress from initiation to alcoholism.

1.3 Rationale of the study

Non-communicable diseases have not been modelled adequately as has been the case for communicable diseases. The effects of NCDs arising mostly from lifestyle choices such as alcohol use are insidious and chronic. Indeed, use of alcohol and drugs continues to In Kenya, just like in many developing countries the emerging pattern of alcohol consumption especially among youth, is worrying due to its negative health and socio-economic impact. Prevailing challenges such as unemployment, neglect, violence, sexual abuse and poor academic performance (NACADA, 2012) have made young people to use alcohol abuse (which leads to alcoholism) coping strategy. By comprehensively investigating risk factors associated with alcoholism using SEM model, modelling incubation period of alcoholism and establishing the hazard of becoming alcoholic then, we will understand better the process of progression to alcoholism.

Developing a mathematical model by taking biophysical process that generated the disease into account, can yield more plausible explanation and interpretation than just fitting model to data. With information on progression to alcoholism sound strategies to arrest the rising alcoholism trend can be put in place and their effects evaluated. Besides models for projecting alcoholism incidences will be vital in planning for resource allocation in the wider health sector.

1.4 Research Objectives.

1.4.1 General objective

To undertake structural analysis of risk factors of alcoholism and model alcoholism as a non-communicable disease.

1.4.2 Specific objectives of the study

1) To analyse the structural relationship between alcoholism and its risk factors.
2) To model incubation period of alcoholism based on frequency of alcohol intake.
3) To evaluate parametrically and non-parametrically the hazard of becoming alcoholic
4) To apply incubation period model in projecting incidences of alcoholism.

1.5 Significance of the Study

The determination of incidences of alcoholism especially in Kenya have mainly been survey based. These surveys focus mostly on prevalence of alcohol consumption. NACADA and recently the Ministry of Health have had renewed interest in establishing the state of alcoholism in the wake of the numerous reported cases of deaths related to alcohol use. The aim is to predict the number of alcoholics in the future in order to guide in policy making on health care systems.

This study added to the existing methods of determining incidences through mathematical model to describe the progression of alcoholism as non-infectious disease. This produced both socio-economic and academic benefits. Models of alcoholism help to understand the nature of the disease in order to mitigate its social and economic burden. For example modelling of risk of becoming alcoholic was critical in quantifying the danger alcoholics face as the frequency of alcohol taking increases. Besides the gender factor in alcoholism can shed more light from a statistical viewpoint

thereby confirming the need to design gender sensitive alcoholism prevention programs. The consumer armed with information on the damage long-term alcohol use could weigh it against any perceived benefit of alcohol use such as the immediate felling of relaxation and then make a rational decision regarding responsible alcohol use. Analysis of risk factors elicits information about which factors are significant hence it can serve as a guide for designing factor based preventive programs.

Ultimately the academia was also to benefit through: innovation in using B-S model in alcohol studies where it hasn't been applied before, introduction of concept of incubation period in alcoholism, study of alcoholism as a non-infectious disease and application of back-calculation method in alcoholism.

1.6 Limitations of the Study

Like any retrospective study, the study of alcoholism face certain challenges which include: Recall error and social desirability since dates of onset of alcohol use (exposure), diagnosis with alcoholism and amount drunk were determined retrospectively hence problems associated with recall are bound to arise. The underlying risk ratios based on recalled information may be underestimated because of their strong reliance on memory; presence of under-reporting, delayed reporting and under-diagnosis affect the most recent alcoholism data series may undermine estimates obtained using back-calculation.

In addition lack of uniformity in terms of volume and alcohol content in the packages. Many alcoholic brands sold in the market differ in concentrations and volume. The frequency of drinking per week and other behaviours in an individual may vary with time. Also patterns of drinking, in particular heavy episodic drinking may be irregular. This study assumed that frequency of drinking per week did not change with time.

CHAPTER TWO: LITERATURE REVIEW

2.0 Introduction

In this chapter the structural relationship between alcoholism and its risk factors, SEM conceptual model and risk factors of alcoholism, models of incubation period of alcoholism, evidence of cumulative damage and the Birnbaum-Saunders distribution, modelling the factors associated with alcoholism, how the risk factors influence the likelihood of alcoholism, hazard rate of alcoholism, hazard function, proportional hazards model in alcoholism, parametric and non-parametric hazard rate of becoming alcoholic, event history analysis, discrete-time models, risk and projecting incidences of alcoholism using back-calculation method were discussed.

2.1 The Structural relationship between alcoholism and its risk factors

The use of Structural Equation Modelling (SEM) has increased in many research fields such as psychology, sociology, education, and economics since its conception by geneticist Sewall Wright in 1918 (Raykov & Marcoulides, 2006). The structural equation model is a tool for analysing multivariate data. It goes beyond ordinary regression models to incorporate multiple independent and dependent variables in addition hypothetical latent constructs that clusters of observed variables might take, making it quite appropriate for theory testing (Bagozzi (1980); Savalei & Bentler (n.d)). Its popularity in social and behavioural sciences has grown tremendously (MacCallum & Austin, 2000). Raykov & Marcoulides (2006) singled out four SEM models found in the literature. A review of development of SEM in research is available in Khine (2013). SEM being a technique for investigating relationships between latent (unobserved) variables and constructs that are measured by manifest (observed) variables or indicators requires the separation of variables into two classes: observed and latent variables, Kline (2011). SEM involve multi-equation regression models that are

multivariate in nature (Fox, 2002). Estimation of relationship between observed and latent variables involves a measurement model while structural model assesses the direction and strength of the relationship between endogenous and exogenous variables. Three characteristics of SEM are:

> A comprehensive statistical approach to testing hypotheses about relations among observed and latent variables (Hoyle, 1995); a methodology for representing, estimating, and testing a theoretical network of (mostly) linear relations between variables (Rigdon, 1998) and it tests hypothesized patterns of directional and non-directional relationships among a set of observed (measured) and unobserved (latent) variables (MacCallum & Austin, 2000)", (Suhr, 2006, p.1).

Study of alcoholism and alcohol use concepts using SEM has been on for some years now. Cooper (1994) utilized SEM to validate a four-factor model on motivation of alcohol use among adolescents in one of the earliest studies applying SEM in study of alcoholism. Chen et al., (2005) evaluated what makes an alcohol advert attractive using SEM. Stamm (2007) applied SEM in testing the relationship between risk taking, injuries, alcohol expectancies, alcohol consequences and alcohol use. Maitso et al., (2008) used SEM to investigate the factors mediating association between first year following alcohol treatment admission and longer term functioning from alcohol use. Green, Polen & Perrin (2009), assessed effect of gender on alcohol use, health and social effects of alcohol use using SEM. Fergusson et al., (2009) applied SEM to test the causal link between alcohol abuse or dependence (AAD) and major depression. Nees et al., (2012) used SEM to study the determinant of early alcohol use in healthy adolescents. They found that reward-related brain activation is likely to fuel addiction more than initiation into early alcohol taking. Moallem et al., (2013) using SEM approach found that greater severity of alcohol use disorder was associated with greater alcohol use.

Recently Kessels & Erreygers (2016) proposed a flexible SEM in study socioeconomic inequality of health. Sznitman & Engel-Yeger (2017), using SEM found that unstructured socializing mediated sensation seeking in adolescent drinking. Li et al. (2017) applied latent structural equation model. SEM was used in identifying risk factors of diabetes by Tripathy et al. (2018) where alcohol use was found to have indirect effect. Kuhn et al., (2019) used SEM to study early brain predictors of development of drinking behaviour between 14 and 19 years. Delgado-Lobete et al., (2020) using SEM, found that alcohol had a mediating effect in university students' tobacco smoking and illegal drug consumption behaviour.

The minimum sample size that must be used in the structural equation modelling method is at least 10 times the number of parameters that can be estimated in the model (Jayaram, Kannan, & Tan, 2004). Some researchers suggest that the sample size for Structural Equation Models should be 200-500, (Celik & Yılmaz, 2013).

Mahugu (*n.d*) found that in Kenya, 77.4% of alcoholics bought alcoholic beverages from bars while 11.4% bought it from wine and spirits shops and 8.6% bought from traditional brew sites. This information was pivotal in designing sample scheme.

Modelling of alcoholism using SEM has not been exhaustively studied.

2.1.1 The SEM Conceptualised Model and Risk factors of alcoholism

SEM accounts for variation and covariation of measured variables. It requires formal specification of model to be estimated and tested. In model specification researchers support hypothesis with theory or research then specify relations a priori while placing very few limitations on types of relations (Suhr, 2006). The main goal on SEM is to explain variance among set of variables and understand the associated pattern of correlation/covariance in the specified model

Kraemer, Kazdin, & Offord, et al. (1997, p. 338) noted that;

> Risk factor is a measurable characterization of each subject in a specified population that precedes the outcome of interest and which can be used to divide the population into two groups (the high-risk and the low-risk groups that comprise the total population)

People initiate alcohol use for different reasons. Pettigrew & Donovan (2003) opined that reasons for continuing to drink may include factors that triggered initiation. ICAP (2009) proposed four broad domains namely genetic predisposition, individual characteristics, social economic factors and environmental determinants. Social marginalisation and economic deprivation modify the relationship between drinking and problems associated with alcohol.

Multiple individual characteristics such as age at onset of alcohol drinking, current age, personality traits, physical and mental health status influence the development of alcohol consumption patterns. For example, Hingson et al., (2006) found that many persons who take their first alcoholic drink at an early age (before 14 years) have a high chance of becoming alcoholic within 10 years of inception. This position is also supported by DeWit et al. (2000) and Gómez et al. (2011) who noted that onset of alcohol use between ages 11–14 greatly increases the risk of development of alcohol use disorders.

The combination of factors that make some individuals become alcoholic and others not despite consuming large quantities of alcohol for years is unknown. "Risk factor is a variable that significantly predicts whether an individual will develop a disorder or disease" (Donovan, 2004, p. 2). Donovan (1977) suggested that the best predictor of abstention or moderate drinking or regular excess drinking was the number and intensity of factors surrounding an individual or group of persons. Thus misuse depends on either more of these factors co-occurring or occurring at a higher intensity.

Research on these factors have mostly been disjointed in that many authors discuss an individual factor or a few factors at a given time and leave the others out. For example, Schuckit (2002) noted that there is solid evidence of impact of genes on alcoholism. Genetic factors may affect alcoholism on their own or combination with other genetically influenced characteristics and often interact with the environmental forces. Risk factors pertinent to an individual are: low self-esteem, aggressive and impulsive behaviour, poor decision making skills and anger management. At community level protective factors include; the participation in making healthy decisions, involvement in community activities, strong sense of religiosity, positive peer relationships and lack of peer approval to alcohol use. While at the family level, protective factors include high levels of family connectedness and family support, opportunities for positive involvement and empowerment as well as a sense of belonging, which promotes healthy decision-making among youth and reduces negative externalized behaviours. Nacoa (n.d) found that "A complex mixture of genetic and environmental factors influences the risk of the development of alcoholism", (p.4).

The larger environment in which drinking occurs significantly affect the patterns and outcome of alcohol consumption. Individual influences cannot be separated from other factors, besides the interactions among different factors are complex (ICAP, 2009).

Models linking alcoholism with risk factors were conceptualised as being multi-level. For example models by Hassan (2013) and ICAP (2009). These conceptualised models were proposed without invoking SEM as seen in the discussions below.

"Lecture 1" (n.d), discusses six-level model of alcoholism where cell was at the lowest level and society (where regulations, organisation and social norms plays the key role) at the highest. Other intermediate levels in order are: organ considering its ability to

metabolize ethanol, the person where the concern is the genetic susceptibility to addiction, family with risk being alcohol abuse at home and neighbourhood where availability of bars is the risk. Interaction between different levels forms a complex matrix. Example family members share unobserved family effect, such as genes, diets, culture, and other unmeasured factors. See Figure G2 in appendix G.

ICAP (2009) grouped risk factors into four broad domains namely; genetic predisposition, individual characteristics, social economic factors and environmental determinants and presented a complex multilevel model shown in Figure G3 in appendix G. They noted that "Research often focuses on these factors one by one; however, their influence is complex and interrelated, and effects cannot easily be disaggregated. Substantial interdisciplinary research is needed to understand better, how different determinants interact" (ICAP, 2009, p. 2).

Hassan (2013) conceptualised a multilevel model for factors related to alcohol use by university students. He suggested that risk of alcohol dependence involves, equally, the environmental and genetic factors, coupled by significant interaction between them. It is represented by Figure G1 in the appendix G.

2.1.2 How the risk factors influence the likelihood of alcoholism

Recent research into the causes of alcoholism emphasizes study of links between biological and psychosocial variables rather than studying each in isolation. Individual characteristics of alcohol consumers such as age, physical and mental health status, stress, and beliefs and expectancies regarding alcohol taking influence the development of drinking patterns.

There is no single, simple explanation for why some individuals develop problems with alcohol. One of the central findings of the large body of research that has examined the

psychosocial causes, or aetiology, of alcohol use is that there are multiple pathways to behaviour that involves alcohol consumption (Cloninger et al. 1996; Sher et al. 1997; Zucker et al. 1994). Multiple biological and psychosocial factors mutually influence each other in causing alcohol abuse; it would be incorrect to view psychosocial causes as either independent from, or competing with, biological causes, ICAP (2009). Rather, alcohol use and alcoholism are best viewed as end products of a combination of biopsychosocial influences, "Chapter 3" (2000). White & Jackson (2004) claimed socio-demographic characteristics or predictors of alcoholism include race or ethnicity, gender, marital status and parenthood, college status and employment.

In study of adolescent risk factor for adult alcohol use and abuse Merline et al. (2008), found that "many adolescent individual and contextual characteristics remain important predictors of adult alcohol use and abuse, and their predictive impact varies as a function of age and type of alcohol outcome", (p.1). Many risk factors of alcoholism are related to drug use (gender, parental attention, socioeconomic status, etc.) perhaps due to exposure to opportunities of taking drugs they create, Danelia, (2011).

Rozin & Zagonel (2012) in a review of 21 articles published between 2000 and 2009 found evidence to show that alcohol initiation occurs mostly between 14 and 16 years. Cox (2007) applied multi-level regression analysis to study factors associated with the age of first use of various substances. Grant & Dawson (1997) determined the odds of alcohol abuse as a function of the age at onset of alcohol use. They found that age at onset of alcohol use is a strong predictor of lifetime alcohol dependence and abuse. Moffitt, (1993); Sampson & Laub, (2003) found that initiation of early alcohol use in adolescence leads to a greater risk of health-related diseases and disorders. Pettigrew & Donovan (2003) opined that reasons for continuing to drink may include factors that triggered initiation.

DeWit et al. (2000) while investigating age at first alcohol use as a risk factor for development of alcohol use disorders found that use of alcohol in early teenage (age 11-14) greatly increases the risk of development of alcohol disorders and related problems later in life. Gómez et al., (2011), confirmed the earlier findings by DeWit et al., (2000), that early use of alcohol heightened the probability of having alcohol problems in adulthood.

Schaaf & Scragg (n.d) identified risk factors for alcohol drinking as demographic, attitudes, lifestyle and feelings. Almasy (2003) while attempting to quantify risk factors as indices of alcoholism, noted that alcohol dependence is a complex disorder involving interaction of numerous genes with one another and with the environmental, cultural and social factors. Mordey, (2015) identified combination of biological factors such as genetics, mental illness and gender and environmental risks and influence such as home and family, age, peer pressure and stress as the main cause of alcohol addiction. Risk factors for dependency are: early onset of use, use by family member, media influence, and troubled relationship with parents, sexual abuse, domestic violence, low self-esteem, curiosity and peer pressure.

Pettigrew & Donovan (2003) opined that reasons for continuing to drink may include factors that triggered initiation with desire for relaxation, assertiveness and self-confidence, enhanced sexuality, the camaraderie, fun and enjoyment, facilitation of social interaction particularly with opposite sex being the key. Other notable social factors include; boredom or lack of things to do (worsened by) unemployment or underemployment, lack of recreational facilities or sporting activities, drinking simply becoming a way of life within peer or family groups. Crundall (1995) provides an overview of the large number of studies on the predictors and factors associated with onset of alcohol consumption. Donovan (1977) opines that the same factors that

facilitate trial, do propel ongoing consumption and also encourage excess drinking ("problem" drinking).

Danelia (2011) compared linear regression, logistic regression and discrete time hazard models in modelling alcohol initiation. By extending the work of Cox (2007) she addressed issue of early onset of alcohol taking and identified gender, SES, family attention, externalizing behaviour and age of first opportunity of using alcohol as factors affecting the onset of alcohol taking.

Different factors may influence different aspects of drinking, such as initial experimentation, later maintenance of regular drinking, and/or the decision to stop drinking. Not only is alcohol use different from alcoholism, but alcoholism itself takes different forms; researchers have suggested that different subtypes of alcoholism may have different aetiologies (Cloninger et al., 1996; Zucker et al., 1996).

The influence of each factor or group of factors on the combinations predisposes one individual to alcoholism and protects another from the same. For example, interaction between genetics and other factors, particularly social and economic variables such as poverty, malnutrition, health status, and drinking culture, influences the likelihood of development of positive (protective effect on the heart) or negative outcomes such as alcoholism

These factors interact in a manner that either enhances positive effects of alcohol consumption or they reinforce the negative outcomes. A good example is the way effects of genetic predisposition are modified by social and economic variables in shaping drinking patterns and outcomes. Multiple individual characteristics such as age at onset of alcohol drinking, current age, personality traits, physical and mental health status influence the development of alcohol consumption patterns. Social

marginalisation and economic deprivation modify the relationship between drinking and problems associated with alcohol.

As seen above risk factors have been analysed in varying combinations. Issues of alcoholism are often multidimensional and complex in nature. Researchers face the challenge of explaining diverse alcohol-related behaviour ranging from simple alcohol experimentation to severe alcohol dependence.

To establish the relative importance of the risk factors of alcoholism SEM was used owing to its advantages. Based on three models discussed in section 2.1.1 two conceptual diagrammatic models of alcoholism given in Figures 3.1 and 3.2 were formed and analysed using SEM.

2.2 Models of incubation period of alcoholism

When applied to non-infectious diseases, incubation period covers both the induction and latent periods as defined by Schoenbach, (2000). This study used the term incubation period to include both induction and latent period. Thus, incubation period is "time from onset of alcohol use to development of symptomatic disease". Incubation period is best described using distribution/model rather than as a single value, (Egan & Hall, 2015).

Venkatesan (2006) noted that incubation period models or survival distribution models are based on non-negative random variables. They include Weibull, Gamma, log-logistic and log-normal distributions among others. These are fitted either parametrically or non-parametrically. The popularity of a model is pegged mainly on its properties. For example Weibull is preferred since it has proportional hazard and accelerated failure time models.

The term incubation period has not been used in relation to alcoholism in literature. This perhaps is due to the way alcoholism has been modelled previously. In fact, the term incubation is mostly used in relation to communicable diseases. As such literature that support the concept of incubation period were highlighted, such as the disease concept of alcoholism and gradual damage to cells due to prolonged alcohol abuse leading to fatigue failure.

Birnbaum-Saunders distribution is useful for describing non-negative data. It is receiving considerable attention, due to its theoretical arguments based on the physics of materials, its properties and its relation to the normal distribution. It is now a natural model of choice in many situations where a quantity exceed a critical threshold due to accumulation of forces, Leiva et al. (2008). Alcoholism is caused mainly by exposure to high levels of blood alcohol for prolonged periods. Alcohol abuse is viciously 'progressive' in its early stages, however alcohol users rarely evolve into addicts in spite of symptoms of abuse (Vaillant, 2003). Factors such as host susceptibility, social environment, stress, mental health, genetic predisposition, age, ethnic group and sex and chance can cause persons exposed to alcohol to experience varying lengths of incubation periods. Therefore, incubation period is best characterized by a distribution. The lognormal, gamma and Weibull are the most frequently considered parametric distributions, in modelling incubation period (Egan & Hall, 2015). However, Exponential and Rayleigh distributions have also been used in analysis of lifetime data (Sarhan & Kundu, 2009). While, Birnbaum-Saunders distribution has, successfully been used in life studies and in material-fatigue life studies.

In 1969, Birnbaum and Sanders developed a fatigue failure model based on physical mechanism of cyclic loadings or stress on a material, which came to be known as Birnbaum-Saunders (B-S) distribution. Although the B-S distribution was originally

proposed as a failure time distribution for fatigue failure under the assumption that the failure is due to the development and growth of a dominant crack by Birnbaum & Saunders (1969), a more general derivation was provided by Desmond (1985) based on a biological model.

B-S distributions have been applied in several non-engineering fields, such as environmental sciences and forestry (Leiva et al., 2008). Podlaski (2008) employed the BS model to describe Diameter at Breast Height (DBH) data for two types of trees, he discovered that the B-S distribution was the model that best described these data, displacing the Weibull distribution.

Dupuis and Mills (1998); Johnson, Kotz & Balakrishnan (1995); Balakrishnan et al. (2011); Ferreira et al. (2012); Leiva et al. (2009); Leiva et al. (2012); Marchant et al. (2011); and Sanhueza et al. (2013) provide detailed applications of B-S distribution. Some extensions and generalization of the B-S distributions are attributed to Bourguignona et al. (2013); Cordeiro & Lemonte (2011); Guiraud, Leiva & Fierro (2009); Sanhueza, Leiva &Balakrishnan (2008). While more recent developments in the B-S distribution can be found in Leiva et al. (2018) and Marchant (2016).

B-S distribution has been used to model fatigue failure, though not in alcoholism. Leiva et al. (2008) noted that argument showing “fatigue" or presence of “cumulative damage" in a mechanism can justify/favour use of B-S distribution. This study utilised the above argument in proposing the use of B-S distribution to model the incubation period of alcoholism. Summary of damage on cells and organs (section 2.1.1) due to excessive use of alcohol provide evidence of “cumulative damage”. The emphasis was on the death or damage to the cellular material by alcohol or its metabolites during the cyclic process of alcohol consumption. The terms failure time and incubation period

has been used to refer to the time between onset of alcohol taking and diagnosis with alcoholism.

2.2.1 Evidence of cumulative damage

Having high levels of toxins in the bodies for long periods of time increases stress on sensitive internal organs and increases the risk of developing long-term health problems. There is a lot of interest in research on effects of alcohol abuse on human cells/ organs in relation to cellular damage, organ failure and development of various alcohol related diseases. Biometrician and physiologist Jellinek (1960), argues that previously, alcoholism was perceived as a legislative, moral and ethical issue; but it is now a medical problem and a subject of wide-ranging scientific study (Ramsden, 2015). Initial studies on alcohol focused on formulating models of the disposition and development of alcohol dependence that integrate both neurobiological and psychosocial findings (Mann et al., 2000).

Alcohol is a potentially addictive psychoactive substance that is absorbed rapidly through the stomach walls and small intestines directly into the blood stream where it is transported to various organs especially those rich in oxygen such as the brain, Dubowski (1985); Alac (2012). It is metabolised mostly in the liver (over 80%), first into acetaldehyde then to acetate. These products are very reactive. The formation of acetaldehyde induces toxic effects by binding to protein and DNA resulting in functional alterations and protein adducts which activate the immune system by forming auto antigens. It also induces mitochondria damage and impairs glutathione function leading to oxidative stress and reduced DNA repair, Palmer (1991). With chronic heavy alcohol consumption, cellular damage as defined by Desmond (1985) occurs.

In addition alcohol significantly alters gut micro-biota (Mann et al., 2000), resulting in an altered balance of pathogenic and commensal organism, ultimately leading to inflammatory and fibrogenic process, Ramsden (2015). Alac (2012) and Rehm et al. (2007) noted that only above average amount of alcohol consumed for a prolonged period of time causes damage. However, there is no safe level of alcohol consumption, Gordis (1997). For some individuals the risk of harm begin to increase at levels below the recommended limits.

The magnitude of injury from alcohol abuse is affected by genetic makeup and environmental factors (Alac, 2012). Rehm et al. (2010a) in a meta-analysis on effect of alcohol consumption on morbidity and mortality found that there exists a dose-response relation between quantity of alcohol taken and magnitude of disease burden. Tuyns et al. (1977) found that those who drunk more than 80grams of alcohol per day had an odds ratio of 5 in getting oesophageal cancer. Rehm et al. (2007) found that the cumulative amount of alcohol consumed is a good predictor of progression to alcoholism. Rehm at al. (2006) opines that frequency of alcohol consumption is a better variable (than quantity of alcohol consumed) in predicting effects of alcohol consumption.

Thus, alcohol use causes damage to the body cells/ organs associated with alcohol metabolism for which Miner's (1945) rule on cumulative damage in fatigue failure was assumed to apply. Smith et al. (1993) showed that the level of BAC spikes shortly after alcohol consumption and then reduces gradually from the body. Mumenthaler et al. (1999), while investigating gender differences in moderate drinking discussed alcohol pharmacokinetics, and presented an alcohol absorption and elimination curve which were similar to those by Smith et al. (1993). They noted that BAC level depends on rate of alcohol absorption from gastrointestinal tract into the bloodstream, the volume of

distribution in the body and rate of elimination. Alcohol absorption and elimination curves given by Dubowski (1985); Smith et al. (1993) and Sprunt (2004) are functions of either zero order or first order kinetic, while Mumenthaler et al., (1999) gives a second order kinetic. Other models of alcohol absorption and elimination in the body show a similar pattern, Sprunt (2004).

Rehm et al. (2010) modelled statistically the volume of alcohol exposure. Rehm (1998b), gave an overview study of alcohol consumption patterns. Bhunu (2012), stated that overall volume of alcohol consumed is the principal underlying factor in alcoholism. Longitudinal studies given by Baer et al. (2001); Fillmore (1988); Vaillant & Hiller-Surmhofel (1996); Vaillant (2003) provide natural history of alcoholism.

More studies on alcohol consumption and its consequences are by Ashley et al.,(2000); Beseler et al. (2008); Bradshaw et al. (2011); Catalano et al., (2001); Chermack & Giancola (1997); Dawson et al. (2005); Frazier et al. (2011)); Gao & Bataller (2011); Gmel & Rehm (2003); Iwanickal & Olajossy (2015); Jernigan (n.d) ; Korhonen (2005); Lim et al. (2012); Merchant (2013); O'shea et al. (2010); Rehm et al. (2003); Rubin (1999); WHO (1999); WHO (2001); WHO (2005); WHO (2014); Wilfred & Day (2007); Woo & O'Brien (2012) and Yi-lang et al. (2013). The above findings provide evidence of cellular damage arising from alcohol abuse, attributed alcohol induced stress on body cells.

Dose-response relationship between alcohol abuse and related harm (cell damage) has not been conclusive as provided in Corrao et al. (1999); Fergusson, Boden & Horwood, (2009); Kucera & Cervinkova (2014); Loring (2014); McKee et al. (2001); Murray et al. (2002); Pollard et al., (2013); Stowell & Stowell (1998) and Read et al., (2003). Low

and moderate amounts of alcohol consumption have been found to have beneficial health effects, (Bellis, 2016; Baliunas et al. 2009; Roerecke & Rehm, 2012).

If alcohol consumption is regular then long-term BAC graph can be assumed to mirror cyclic stress pattern used in deriving B-S distribution by Desmond (1985), which has not been considered in any of the previous alcohol studies.

2.2.2 Cellular damage and the B-S distribution.

Regular drinking cause depletion of the neurotransmitter serotonin leading to a cyclic process of drinking to relieve depression as levels of serotonin become more depleted, (Cornah, 2006, p. 7). The psychological aspect of alcoholism (dependence) is related to depletion of neurotransmitters (the brain's 'messengers') needed to reduce anxiety naturally.

> Keeping the optimum balance of alcohol to reduce anxiety is almost impossible because the effect of alcohol on the brain is such that after the initial 'euphoria' or stimulation from the first drink, alcohol acts as a depressant and the feelings of anxiety may rapidly return. (Cornah, 2006, p. 7)

Increased drinking to cope with those feelings leads to a rapid increase in the levels of alcohol in the blood leading to a cyclic course of depletion, more anxiety or depression, needing more alcohol to cope, higher BAC and more depletion of neurotransmitters. Two conclusions that can be deduced from the above observation is that: first is that this cyclic process mirror the process that was used in the derivation of B-S distribution. And secondly the frequency of alcohol consumption can be a good predictor of time from onset of alcohol use to alcoholism.

From a theoretical point of view evidence of cellular damage arising from a cyclic pattern of alcohol abuse provide strong basis for linking B-S distribution to alcoholism using results of Desmond (1985).

Literature describing use of Birnbaum-Saunders distribution in modelling incubation period was unavailable by the time of this write up. B-S distribution was suitable for modelling incubation period of alcoholism because of the theory behind progression to alcoholism and the fitting of the empirical data.

2.3 The hazard of becoming alcoholic

In this section we reviewed hazard function, proportional hazards model, Parametric and non-parametric approaches used in finding hazard rate, event history models and discrete time models. Literature on risk of becoming alcoholic was also reviewed.

2.3.1 The Hazard function

This function shows the dependence of chance of failure (in our case onset of alcoholism) on time. Leiva et al. (2008), suggested that the shape of the hazard function arising from say analysed data on frequency of alcohol intake can be used to assess or confirm the suitability of the B-S distribution in describing alcoholism. Rundel (2012) opined that hazard rate can uniquely identify a distribution.

Several articles have been written on the B-S distribution and its properties in the last four decades. However the shape of the hazard function remains unexplored perhaps because of its complex form, Leiva et al., (2008). Mann et al.(1974) claimed without proving, that hazard function of the BS distribution is a non-increasing function of t. B-S distribution has upside-down hazard function of $t>0$ for all values of shape and scale parameters (Leiva et al., 2008).

Obtaining an upside down shape of the hazard was a confirmation that the data was best modelled using B-S distribution.

2.3.2 Proportional hazards model and alcoholism

Hazard functions display the instantaneous rate of death (or failure) assuming that the subject has survived up to that time. Intuitively, it is an indicator of the risk occurrence of the event for an individual at a given point in time. There is need to assess the impact of risk factors on the likelihood of becoming alcoholic using proportional hazards model. Proportional hazard model is suitable in connecting time from onset to diagnosis to some covariates. It is very popular in modelling time to event. It can also be used to determine how risk factors are associated with onset of alcoholism. As noted earlier, most studies focus on one factor at a time for example; Xu (2004) reviewed application of Cox proportional hazards mixed model on effect of genes on alcohol use in twins by I-Chao Liu (2004). Nicholson et al. (2005) determined Cox proportional hazard ratios for the effect of relatives' characteristics on risk of death from all causes in a study on alcohol consumption and increased mortality in Russia; Wagner et al. (2005) used Cox model for discrete-time analysis with stratification to estimate the risk of drug use associated with early alcohol and tobacco initiation; Staveren et al. (2006), used Cox proportional hazards model to assess the effect of diet scores and different components with alcohol being present or absent in the diet; Hingson et al. (2006) used Cox proportional hazards multiple regression models "to assess increased risk of dependence; Xu et al. (2009) applied proportional hazards models in selecting models using profile likelihood for semiparametric models.

In 2010, Holahan et al. estimated the effect of daily alcohol consumption on mortality risk, using Cox proportional hazards regression models. Howie et al. (2011) used Cox proportional hazards regression models to determine the hazard ratios in a study of alcohol consumption and risk of all-cause and cardiovascular disease mortality in men. Shuval et al. (2012), applied Cox proportional hazard models to evaluate how exposure

variables (which included alcohol) and incidence of Metabolic Syndrome in men were related.

Using Cox proportional hazard regression model Midgette & Kilmer (2015) modelled both the time to violation and the risk of re-arrest in a study effect of Montana's 24/7 sobriety program on driving under the influence re-arrest. Sjolund (2015) used Cox proportional hazards regression models to model association between IQ and alcohol-related admission to hospital and death. Koning (2015), used design-based Cox proportional hazards regression models to address oversampling of subjects in a study on association between alcohol consumption and the risk of developing chronic kidney disease instead of using stratum-specific baseline hazard functions. Keller (2016) used piecewise-constant proportional hazard model in longitudinal analysis of relationship between alcohol and financial success. Borges et al. (2016) used Cox regression model to estimate the likelihood of healing of wound with alcohol consumption as a variable. Whitman et al. (2017), applied Cox models to investigate associations and independence of effects of alcohol abuse and estimate cumulative incidence of three cardiac diseases. Gowin et al. (2017), used Cox proportional hazards models and concluded that binge drinking may be an early indicator of vulnerability to alcohol use disorder; Canchola et al. (*n.d*) investigated the alcohol use as a risk factor in breast cancer using Cox regression models.

Humphreys et al. (2017) used Cox proportional hazard models to test for association between alcohol taking and transaminitis singly, then adjusted for age and gender. Katikireddi et al. (2017), used Cox proportional hazards models to investigate associations between exposures of interest and the first episode of an alcohol attributable outcome. Clearly the factors related to alcoholism and alcohol use are mostly studied singly. Cox proportional hazards model was used to evaluate the effect

of multiple risk factors of alcoholism on the hazard function. It is a product of two components; time and risk factors. One of its component depends on time while the exponential part that depends on the other risk factors. The model assumes that, on log scale, a change in a risk factor results in a proportional change of the hazard rate (Dey, 2020).

2.3.3 Parametric and non-parametric hazard rate

Literature on hazard rate spans both in engineering and biological systems. Term hazard rate is commonly used in biostatistics (means that the event is harmful), Danelia (2011). Hazard rate can be determined using parametric or non-parametric methods. Yamaguchi (1991), gave definition of hazard rate of initiation.

Hazard of onset of alcohol taking was determined by Danelia (2011). She used discrete time hazard models to estimate the risk of onset of alcohol taking in adolescents.

2.3.4 Event history analysis.

Event analysis is the study of the duration of non-occurrence of an event during the risk period, Yamaguchi, (1991). Although terms survival analysis and event history analysis are often used interchangeably, the term event history analysis is used mainly in social science, to refer to events that are repeatable and an individual's *history* of events is of interest. In our study the interest is how covariates affect "duration or incubation period". That is how long normal alcohol taking persists before culminating into alcoholism. Events, such as the occurrence of a disorder like alcoholism represent a transition from one state to another. The event of concern is tied to the history (factors and time) preceding that event. The dependent variable measures the duration of time that units spend in a state before experiencing some event. Event history is longitudinal

and involves statistical analysis of longitudinal data collected on a set of observations, Box-Steffensmeier & Jones (2004).

Baer (2005), cites a sample of recent event history models in social sciences and other fields. He noted that event history models are often used in describing transitions of events such as in criminology where the event of interest is recidivism; in health event of interest may be say transition from HIV positive to having AIDS symptoms.

Event history data are often collected retrospectively. Respondents are asked to recall the dates of events that have occurred since a certain event took place such as onset of alcohol taking. “Respondents may be asked to recall events in the order that they occurred or in reverse chronological order, depending on the significance of the start of the observation period with reference to the process under study”, (Steele, 2005, p.4). Washbrook and Steele, (2013) describes discrete-time event history analysis with examples on both fixed time covariates and varying time covariates.

Event history analysis makes it possible to estimate time periods the event of interest is most likely to happen, and to determine why some individuals experience the event earlier than others and why some do not experience the event of interest at all during the study period, Tekle & Vermunt (2012). Even though event history analysis helps to answer such questions, it also poses certain challenges that are hard to deal with using standard data analysis techniques such as linear and logistic regression analysis (Allison, 1982; Tuma & Hannan, 1979; Willett & Singer, 1993). Thus simple linear and logistic regression methods are unsuitable for dealing with two unique features of event history data; i.e. censoring and time-varying covariates. Therefore, special regression techniques known as event history models, hazard models, survival models, failure time models, and duration models are needed, Tekle & Vermunt (2012).

Survival analysis or event history analysis is usually used when the research revokes the test of "whether and when", Singer and Willet (2003). This study passes this test since it investigated whether or not alcohol disorder occurred and when it did (how many years have passed from onset of alcohol taking to diagnoses with alcohol use disorder). Petersen (1995) noted that event history data are obtained from failure time processes. At any given point a unit in the process, is "at risk" of experiencing the event. Where an event represents a change or transition from one state to another, Box-Steffensmeier & Jones (2004).

Event history models periods of time or duration during which respondents are "at risk". In the current study an individual becomes "at risk" once he/she starts using alcohol regularly up to the time he/she is diagnosed as being alcoholic. Event was an indicator of transition from normal/moderate alcohol taking state to state of diagnosis with alcohol use disorder.

Censoring occurs whenever an observation's full event history is unobserved. Singer and Willett (2003), opine that validity of hazard analysis is based on the assumption that censoring is non-informative. If at the time of interview a person is not yet alcoholic he/she is censored. Right censoring was used because the duration of time until alcoholism is not known since the event occurrence (onset of alcoholism) has not been observed. Censoring occurred at the same point in time for all individuals (age at interview date). According to Neels (2014), if censoring is non-informative risk set can be assumed to represent all individuals who would have been at risk of event occurrence if everyone could have been followed that long. Besides under the assumption of non-informative censoring the experience of each interval's risk set can be generalized back to the entire population.

Hoffmann & Cerbone (2002) explored the risk associated with adolescent drug abuse and parental substance use disorder using event history analysis. They found that unobserved physiological sensitivities put the children at particular risk to alcohol abuse. This was followed by Crawford et al. in 2009 who found that adolescents that had substance use disorders were more probable to engage in violence. Later Borges et al., (2011) undertook a cross-national study using discrete-time event history models to study prevalence of substance use and substance use disorders among Hispanics in the US, taking into account the time-invariant and time-varying characteristics.

The relationship between socio-economic status (SES) of parents and substance use by their adolescents using event history analysis was examined by Sutherland (2012). King (2014) used Bayesian event history analysis to model recurrent episodes of illicit drug use. Yang et al. (2017) used event-history analysis with risk free model to characterize alcoholism susceptibility and age at onset simultaneously. More recently, Maggs et al (2019), used event history models to predict onset of alcohol use for early drinking initiation, while Mou and Lin (2020) used event history methods in the study of onset of alcohol use by Chinese college students. They found that drinking as preparation for adulthood was a strong predictor of onset of alcohol use.

2.3.5 Discrete-time models

Often, researchers in substance use disorders examine critical events such as onset, recovery from illness and relapse. Data on such events are collected say weekly or monthly then discrete-time survival models fit better than continuous-time models, Xie et al. (2003).

Time is a continuous variable since changes of state may occur at any time. Event History models used nowadays mostly involve “continuous time”. However, durations

are finite (e.g., months, weeks, years). Discrete data usually introduce ties, hence the need to use true discrete-data models, Broström (2012). Discrete-time model examines length of time to move from one state to another (interval between consecutive changes of state defined by some qualitative variable within some observation period). Essentially, discrete-time models are logistic regression models, Steele (2005).

Discrete-time models have many advantages over continuous-time models. The main disadvantage is the need to restructure the data prior to analysis. In a discrete-time event history data set, each individual contributes multiple records. Repeated measures on the ith respondent may exhibit either duration dependency or temporal dependence, Box-Steffensmeier & Jones (2004). The first step in a discrete-time analysis is to create person-period data set. . Hence each individual's record is replicated as many times as the observed number of time intervals until either the event of interest or censoring occurs, (Stewart, 2010). Using of discrete-time models Xie et al. (2003) found that gender and age were strong predictors of remission in persons with severe mental illness.

2.3.6 The Risk

Finkelstein (2008) found that one can use information on the process of a 'failure development' in modelling failure rate. Such as when failure occurs due to accumulated random damage or wear exceeds a predetermined level. Here the failure rate can be derived analytically considering stochastic processes of wear. An individual who is taking alcohol normally has a risk/probability that he/she will develop alcohol use disorder. This probability of becoming alcoholic in the time interval [0, t) is given by Gail (2005) as;

$$1 - \exp\left(-\int_0^t h(u)du\right) \qquad 2.1$$

Assuming that the frequency of alcohol taking per week is constant for a given alcohol user, and that it does not change throughout the alcohol taking life, then risk of an individual who takes alcohol *t* times a week being diagnosed with AUD is given by equation 2.1. Where, $h(u)$ is the hazard function.

2.4 Projections of incidences of alcoholism

In this section mathematical models of alcoholism, back-calculation method, and surveys of alcohol use and alcoholism in Kenya are discussed.

2.4.1 Mathematical models of alcoholism

The study of epidemiology of alcoholism using mathematical models started after the emergence of disease concept of alcoholism. Most of alcoholism models assume infectious disease concept as given by Kermack and McKendrick in 1927 (Chowell et al., 2009). It considers a constant population where individuals are split into compartments of those that are susceptible to catching the disease (S), infected individuals (I) and immune or dead individuals (R). Deterministic or stochastic models are useful in describing the mechanism of the infection, Ahrens & Pigeot (2006).

Several authors have modelled different aspects of alcoholism for example Field (1985) studied statistical distribution of alcohol consumption and consequent inferential problems while Skog (1985) analysed collectivity of drinking cultures. Mathematical modelling of alcoholism in normal and chronic alcohol users has been analysed by Smith et al., (1993). With Shirley et al. (2010) using hidden Markov models to study alcoholism treatment.

In a 2007 PhD dissertation, Sanchez (2007), presented a ground breaking paper entitled “Studies in epidemiology and social dynamics” whose second part contained one of the most influential infectious disease model on alcoholism, entitled “modelling dynamics

of harmful alcohol use". It involved use of ordinary differential equations to model a system of interaction between three epidemiological states; susceptible, drinkers and recovered. Later in the year (Sanchez et al., 2007) published a paper on drinking as an epidemic using simple mathematical model with recovery and relapse. Benedict (2007) showed how "infected buddies spread the problem drinking". Cintron-Arias et al. (2009) introduced the "Role of non-linear Relapse on contagion amongst drinking communities", while Manthey et al. (2008) studied campus drinking using an epidemiological model. Teymuroglu (2009) gave continuum models for spread of alcoholism. In 2013 Walters et al., modelled alcohol problems having recovery option while Bhunu (2012) analysed alcoholism using mathematical model. Sharma and Samanta (2013) discussed drinking as an epidemic incorporating dynamic behaviour in the mathematical model. Brennan et al (2008) did "Modelling the potential impact of pricing and promotion policies for alcohol in England".

These papers by Benedict, Bhunu, Manthey and Sharma & Samanta, took the approach proposed earlier by Sanchez in 2007 with a slight variation to account for difference in each situation. Their presentations were founded on the idea that the outcomes (patterns) associated with various biological and sociological processes (behaviours) are often the result of interactions or contacts between individuals, groups, subpopulations, or populations. The focus being the 'spread' of alcoholism in a system.

Challenges arise from the fact that the dynamics of these processes at the population level are the result of nonlinear interactions between individuals in different states such as Susceptible (S), Infected (I) and Recovered (R), Sánchez et al. (2007). Where susceptible includes both the non-drinkers and social-drinkers. The alcoholic models described above, rely on application of differential equations. These deterministic

models describe system consisting of the three alcoholism epidemiological states described earlier.

More contributions to the use of differential equations in deriving infectious alcoholism model are available in French et al. (2010) and Cintron-Arias et al. (2009).

2.4.2 Back-calculation Method

There are several mathematical and statistical methods available for estimating incidence patterns and trends. These include; dynamical models, demographic models, back-calculation techniques and birth cohort methods, UNAIDS (2010). Indirect methodologies such as back-calculation developed in the context of HIV have contributed to the development of statistical techniques that can be applied to other diseases.

Back-calculation method has been applied in study of AIDS, which exhibits a relatively long incubation time. It reconstructs the pattern of past infections and predicts the future number of cases using the present infection status, Venkatesan (2006). This method has been useful in generating incidence curves, determination of prevalence and making projections of infectious diseases. Incidence data has been used to reconstruct incidence trends using statistical methodology that was originally developed specifically to deal with AIDS surveillance data, Becker & Marschner, (2001).

Brookmeyer and Damiano (1989) refined back-calculation method in their quest to develop an AIDS model under constraint of inadequate data. Back-calculation method has also been used in non-disease cases, Ward et al., (1989); Holtby et al., (1990) and Fukuwaka (1996). In late 1990s, back-calculation technique was applied during the bovine spongiform encephalopathy (BSE) and variant Creutzfeldt–Jakob disease (vCJD) epidemics. The two diseases exhibit long incubation periods. Susceptibility of

a person to infection was found to be age-dependent, (Egan & Hall, 2015). Age structure was hence incorporated into the back-calculation process for these particular applications.

Earlier in this century back-calculation was used on bioterrorism cases in US anthrax attacks in 2001. Retrospective analysis of the outbreaks enabled estimation of the number of cases that might be prevented by the distribution of antibiotics to potentially exposed persons, Brookmeyer, Johnson & Bollinger (2003), as cited in Egan & Hall, (2015).

Back-calculation method is popular because it requires few assumptions and parameter inputs and has efficient application compared to other modelling approaches. It enables unobservable features of an event (such as onset of alcohol drinking) to be inferred after the occurrence (outbreak).

There are several versions of back-calculation method. Several authors have documented comprehensive reviews of the back-calculation technique. These include Eze (2009), who highlighted strengths, weaknesses, evolution and modifications; Egan & Hall (2015) who reviewed the Back-calculation methods with a special focus on non-transmissible infectious diseases. Globally back-calculation method has been applied in alcoholism in forensic studies of alcoholics involved in traffic accidents.

Back-calculation technique requires the knowledge of information about the time it takes to progress from the onset of alcohol taking to being diagnosed as alcoholic and reported incidences of the disease. This is applied in deriving incidence information on the onset of alcohol taking (often unknown). The desire is to project future cases of alcoholism considering that is only a small proportion of alcoholism cases are reported with majority of alcohol users not experiencing alcoholism while another portion is

known to be abusing alcohol but carry on with their life activities normally and don't seek medical help over alcohol related illnesses (they are called functional alcoholics).

The model takes the following general form

$$Alcoholism\ diagnosis\ date\ =\ Alcohol\ use\ onset\ date\ +\ Incubation\ period$$

Information on time *d* incubation period in form of probability function $f_t(d)$, and the sequence of diagnoses counts of alcoholism , $\boldsymbol{Y} = \{Y_t\}$ for $t\ = 1,2, \ldots T$ are the basis of back-calculation. Where t is the time. Now, incubation distribution and alcoholism incidence data can be used to obtain the unobserved sequence of alcohol initiates $\boldsymbol{H} = \{H_t\}$. Assuming that , $\boldsymbol{Y}$ is a vector of Poisson random variables with mean $\boldsymbol{\mu}$ one can estimate incidences of initiations into new alcohol users, $\boldsymbol{\lambda} = E(\boldsymbol{H})$ by applying the discrete version of Brookmeyer & Damiano (1989) equation connecting the three quantities.

$$\mu_t = \sum_{x=1}^{t} \lambda_x f_x(t-x) \qquad 2.2$$

(Becker & Marschner, 2001)

Where: λ_x is the mean of alcohol initiates incidence at time *x* also called exposure curve, $f_x(t-x)$ is the probability density function for someone who initiated alcohol use at time *x* and diagnosed at time *t.*

The matrix format of equation (2.2) is equation (2.3).

$$\boldsymbol{\mu} = \boldsymbol{X\lambda} \qquad 2.3$$

Where $\boldsymbol{X}$ is the matrix representing probabilities of transition from one period to another (design matrix). $\boldsymbol{\lambda}$ can now be determined using deconvolution or suitable matrix operation. Alternative methods of estimating $\boldsymbol{\lambda}$, are available, Jewell, (1990).

The exposure curve λ_x, is either parametric or non-parametric. Its choice has consequences on estimates projected. In parametric back-projection, a particular functional form is assumed for the HIV incidence curve or the AIDS incidence curve. Strong parametric models such as exponential generate estimates that fit distant past well but fit poorly in the recent past. They include use of step functions, splines or other form of parametric function, Eze (2009). Flexible models such as step functions were proposed by Brookmeyer & Gail (1986, 1988). Rosenberg & Gail (1991) suggested the use of spline functions to overcome the problem of discontinuities in step functions.

Artzrouni (2004) modelled cumulative HIV infections using logistic function. Isham (1989) used a strong parametric function to AIDS incident data then solved for infection curve using equation (2.2). Rosenberg and Gail (1990) proposed a weakly parametric approach which uses splines for smoothing. One advantage of nonparametric models is that they allow data to speak for themselves in determining which configuration best suits the observed counts. Becker et al. (1991) evaluated infection curve using nonparametric method then smoothed the estimates by applying weighted moving average at each iteration in EM algorithm used to maximise the likelihood.

Those who use nonparametric back-calculation method often apply smoothed non-parametric estimates of infection curve as a guide to a suitable parametric model, (Becker 1990). One drawback when incorporating incubation period distribution in back-calculation method is the common assumptions that the incubation distribution is stationary throughout the period and that the probability of progression to alcoholism is the same for all individuals consuming alcohol. These two are highly contestable. Studies have shown that progression to alcoholism depends highly on genetic makeup and age of onset of alcohol taking.

Often alcohol consumption is a not a one off activity but rather continuous process, in which an individual drinks on regular basis. During a drinking session alcohol consumers often take several sips. This may last a few minutes or several hours. Therefore, the person is repeatedly exposed to the disease causing agent over an extended period of time. Hence exposure times can suitably be modelled as a function $\lambda_x(t, \boldsymbol{\beta})$ where, t is time-dependent variable and $\boldsymbol{\beta}$ a vector of parameters, Egan & Hall (2015). While the pdf of times of diagnoses (assumed to be onset times for alcoholism) is given by equation (2.4).

$$a(t; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \int_{T_0}^{T_L} f(t - s; \boldsymbol{\alpha}) \lambda_x(t; \boldsymbol{\beta})\, dx \qquad 2.4$$

Where, $f(t - x; \boldsymbol{\alpha})$ is the incubation period distribution, represented by B-S distribution

B-S distribution was used to model alcoholism incubation period based on biophysical processes, theory behind progression to alcoholism and the fitting of the empirical data.

Equation (2.3) which was established by Brookmeyer and Gail (1988), also applies if the functions $a(t; \boldsymbol{\alpha}, \boldsymbol{\beta})$and $\lambda_x(t; \boldsymbol{\beta})$ represent intensities of the point processes such as annual number of diagnosed alcoholics and exposures, respectively, rather than probability densities (Brookmeyer, 1996).

2.4.3 Survey of alcoholism in Kenya

In Kenya, several surveys on alcohol use have been carried out. In 2010 a baseline survey done by NACADA on alcohol use in Central Province of Kenya) indicated that there is a relatively high level of alcohol use in that province, NACADA (2010). There was fear about the penetration of second generation alcohol. Lilleskov & Chakua (2013) found the prevalence of alcohol use in private sector to be 65.9% and 57.9% in public

sector. NACADA (*n.d.*) put it at 23.4% among secondary school students. Hassan (2013) found the prevalence of alcohol use among University of Nairobi students to be 63.2%. Baridi (2014) studied strategies of regulating alcoholism and other drugs in Kenya. Maithya et al., (2015), found the prevalence of alcohol use was 47% among students in technical institutes. Kaithuru & Stephen (2015), investigated the impact of alcoholism on work force and found a strong relationship between alcoholism, workplace unproductivity, financial problems, stress, hangovers, and diseases which hamper efficiency at Kenya Meteorological station in Nairobi. Incidentally most of these surveys combine the study of alcohol with study of drug use. Alcohol use is different from alcohol abuse, therefore having separate information on the two gives a more realistic picture of the trend in the study of alcoholism. This will help create the trend curve for incidences of alcoholism.

Jewell (1990), suggested inclusion of specific information in the back-calculation method where additional information may be available. This study has taken this cue by first focussing on suitability of B-S distribution in modelling incubation period of alcoholism.

Whereas the concept of damage to organs by excessive alcohol use has been discussed in several researches the current study has not encountered literature linking this damage to B-S distribution. Time from onset of risky alcohol taking to diagnosis with alcoholism was conceptualized and modelled using the B-S distribution. This study evaluated the hazard rate of alcoholism using both the parametric and non-parametric methods. Back-calculation method was used to model alcoholism progression as a non-transmissible disease (WHO, 2000), by applying incubation period determined using B-S distribution.

This study has not found literature on study of alcohol consumption and alcoholism using the process and modelling as a technique described above by the time of compilation of this research.

CHAPTER THREE: METHODOLOGY

3.0 Introduction.

In this chapter method used in data collection, modelling time from onset of alcohol taking to diagnosis with alcoholism and the associated theoretical foundation, structural equation model of onset of alcoholism, hazard of becoming alcoholic and back-calculation method were discussed.

3.0.1 The sample.

Both primary and secondary data were used. For primary data, the sample was taken from seven counties in Kenya, purposively selected based on information from NACADA regarding alcohol use and abuse, their economic, education, ethnic and cultural diversity.

A cross-sectional study which applied multistage stratified sampling method was then conducted in which retrospective self-report, obtained from interviews and survey questionnaire administered to respondents between 6th October 2018 and 30th April 2019. A research permit was obtained from NACOSTI with each county sampled approving the study. Consent was also obtained from potential respondents prior to commencement of the interview. The sample consisted of 780 (see APPENDIX D)

Questionnaire was piloted on current alcohol users from four sub-locations (two urban and two rural) and alcoholics from three rehabilitation centres between 4th and 29th August 2018 in Nakuru and Kiambu counties. Cronbach's alpha was used to assess reliability of the instrument.

These seven counties; Nairobi city, Nakuru, Uasin Gishu, Kirinyaga, Nyandarua, Kiambu, Nyeri had about 25% of the country's population (based on 2019 census report), and were assumed to reflect Kenya's population in terms of distribution of rural

and urban areas, demographic features and socio-economic status, educational, income levels, broad range of cultures, economic development and diversity in addition they are the most affected regions as indicated by the number of alcohol selling points.

Secondary data used was provided by Ministry of Health. It consisted of reported cases or incidences of morbidity and mortality resulting from alcohol use for the period starting from 2014 to September 2019. Incidences of morbidity consists of aggregates for all categories of alcoholism (see Appendix A).

3.0.2 The Questionnaire.

The questionnaire (see appendix E) was intended to assess nine major risk factors of alcoholism; gender, personality, peer influence, family and family attention, environment or structural settings, drinking habits and patterns, socio-cultural, economic status, age at onset of drinking. Interview was used to assess whether the participant had AUD using AUDIT and also get detailed information on frequency of drinking. Interviews were audio-recorded. The inclusion criteria used was; persons currently using alcoholic beverages or persons in rehabs suffering from alcoholism. Participants in rehabs were not subjected to AUDIT, only their drinking habit before admission was enquired. Those in rehabs suffering from alcohol and drug abuse were excluded.

3.1 How Structural Relationship between Alcoholism and its Risk Factors was Analysed

In order to undertake structural analysis of risk factors of alcoholism some key steps were followed. They included; Review of factors of alcoholism for model specification and identification, data collection and statistical analysis which involved scaling,

parameter estimation, assessment of model fit, presentation and interpretation of results. They are discussed below.

Based on literature on factors influencing continued alcohol use highlighted in sections 2.1 and 2.3, nine factors were selected. These were:

- Personality or individual characteristics which includes; personal traits, physical and mental health which either protects or propels one into sustained alcohol drinking.
- Peer influence; this regards peer interaction in environments that promotes desire for alcohol taking.
- Economic Status which includes; economic deprivation indicated by residential area, education level, housing condition etc. that exposes an individuals to risk of alcohol use.
- Social-cultural; social marginalisation and social network, practices and norms, religious beliefs. These can be protective or promotive with respect to alcohol use.
- Environment or structural setting which includes; alcohol laws, drinking culture, availability of alcohol and saliency. These provides suitable or unsuitable context for the other factors.
- Family and familiar attention; the habit running in the family tree. These variables are conceptualised through questions in the questionnaire regarding having alcoholics in the family tree. One may be genetically prone to alcoholism.
- Age at onset of alcohol consumption
- Gender.
- Drinking habits and patterns

The two diagrammatic models that were conceptualised using these nine factors and based on discussions advanced in Section 2.1 of the literature review, are given in Figure 3.1 and Figure 3.2. The study opted to use Figure 3.1 which is a hybrid of arguments by ICAP's, (2009) and Hassan's, (2013) multilevel models. SEM equations were then formed based on Figure 3.1.

Owing to the large number of permutations of variables only two types of relationships mentioned below were evaluated and discussed i.e. Association, e.g., correlation, covariance and Direct effect (directional relation between two variables), e.g., independent and dependent variables. The indirect effect (the effect of an independent variable on a dependent variable through one or more mediating or intervening variables) was not evaluated.

The measurable variables were classified as dependent or independent. Then items were purposively selected to correlate to an external criteria (construct they measure) and not necessarily with each other. The independent variables were first subjected to exploratory data analysis. Then, Factor analysis including drawing a scree plot, cluster analysis and finally SEM were applied on the data.

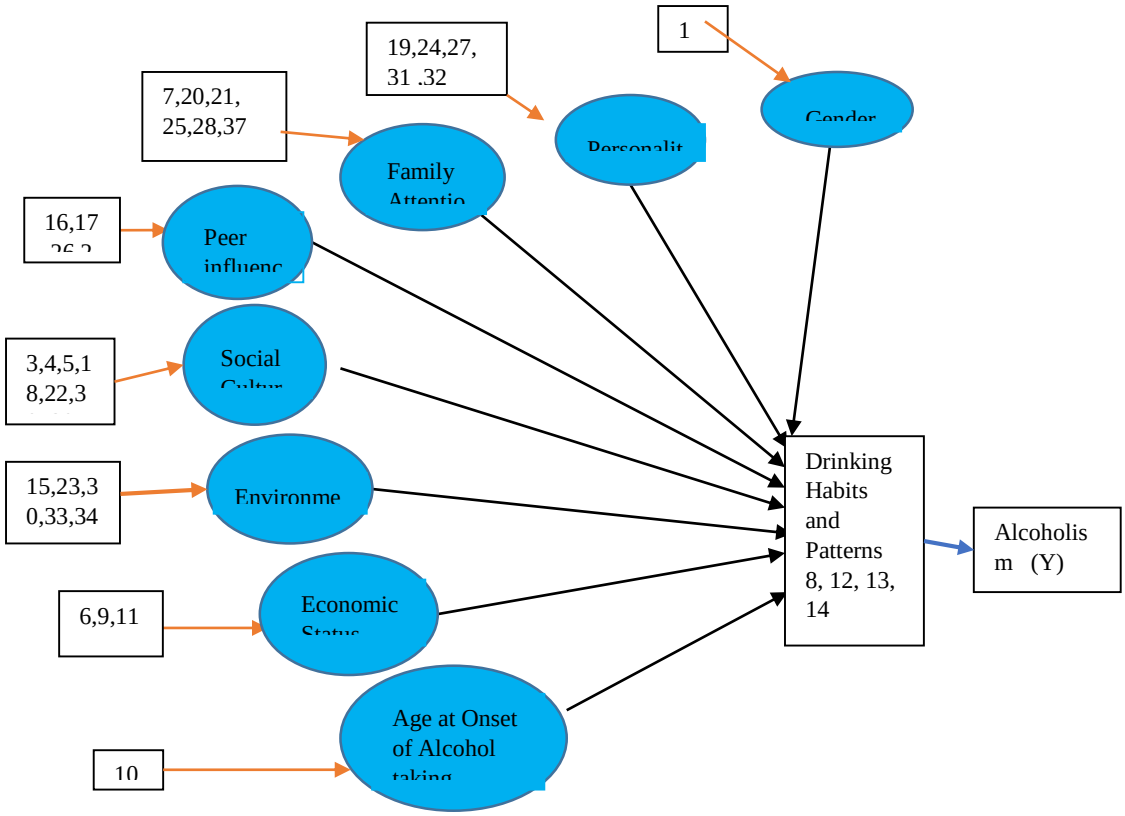


Figure 3.1: Hierarchical model with drinking choices and patterns determining alcoholism.

Principal component analysis (PCA), was done to determine scale validity, make the dataset analysable, remove dependency between variables and to obtain fewer new variables which are not related to each other, (Aksu, Eser, & Guzeller, 2017).

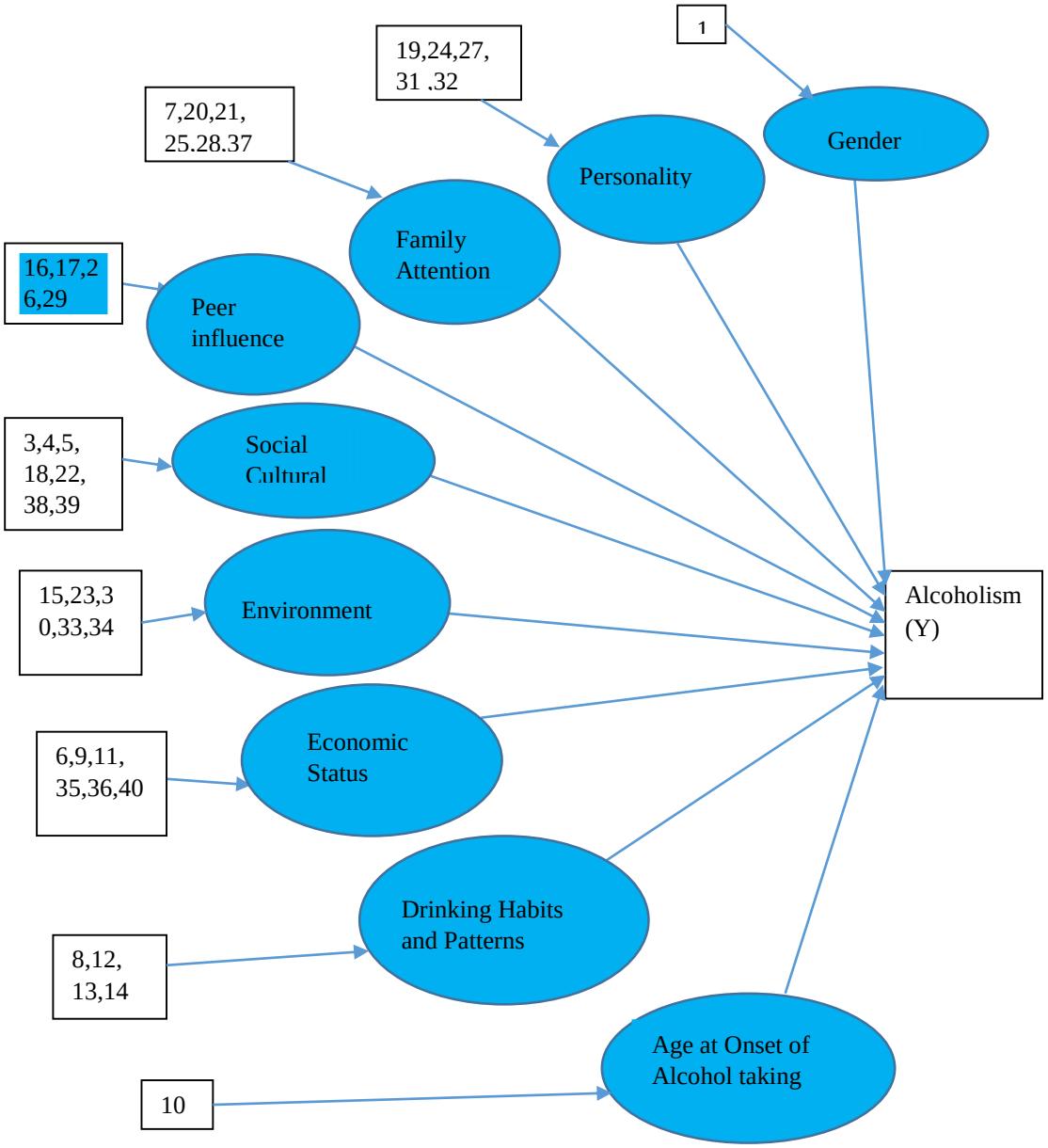


Figure 3.2: A conceptual SEM model of all factors influencing alcoholism directly

Note: The numbers in the boxes represent the item number of the measured variable as it appears in the questionnaire.

3.2 Modelling Incubation Period of Alcoholism

To model incubation period of alcoholism, Birnbaum and Saunders distribution was used. In order to show the suitability of B-S distribution in modelling incubation period of alcoholism, both theoretical and empirical approaches were applied.

3.2.1 Theoretical approach

First theoretical foundations for suitability of B-S distribution were sought from literature. The use of theoretical foundation for applying the B-S distribution instead of other life/ failure time distributions in modelling incubation period of alcoholism is based on the recommendations by Leiva et al. (2008). They opined that argument favouring “fatigue" or “cumulative damage" is a strong reason for invoking Birnbaum-Saunders distribution in describing lifetime data. Three theoretical reasons that support the suitability of B-S distribution in modelling incubation period of alcoholism were obtained from literature.

3.2.2 Empirical approach

Secondly B-S model was fitted to the empirical data obtained using questionnaire and interview, then goodness of fit assessed. In order to account for possible multiple drinking occasions per day the study defined a drinking session as a continuous and uninterrupted drinking spell separated from the next session by a break of at least four hours (enough duration for BAC to drop appreciably). This was occasioned by a number of individuals who take alcoholic drinks several times in a day (in the morning, afternoon and evening, with long breaks in between).

The results of frequency of drinking during the week preceding interview for the respondents found to have AUD were used in modelling B-S distribution. They were summarised in Table 4.2. For those in rehabs the data relates to the week preceding

admission. The data was first subjected to exploratory data analysis. Owing to difficulty in finding explicit expressions for the estimators in B-S distribution, numerical procedures are used (Leiva, 2016). Three methods; Maximum Likelihood Estimation, Modified Moment Estimation and graphical methods were used to estimate the B-S model parameters $\alpha \; and \; \beta$.

The study assumed that effects of other drugs abused alongside alcohol, which is very common had no significant confounding effects on alcohol effect and therefore did not impact on incubation period distribution

Using these values of the MLE's for alpha and beta, model was fitted to data. TTT plot, p-p plot, Q-Q plots; r-squared and Kolmogorov-Smirnov tests were done to assess goodness of fit for the model. Hazard function was also evaluated to establish whether it was characteristic of B-S distribution. To implement the above operations bs, VaRES and VGAM packages in R-Studio Version 1.2.1335 were applied.

3.3 Evaluating Parametric and Non-Parametric Hazard of Becoming Alcoholic.

The hazard rate of becoming alcoholic was determined using parametric and non-parametric methods. The parametric hazard rate was modelled using hazard function of B-S distribution as given by Kundu et al., (2008). MLE for the parameters α, β were evaluated by setting both $\alpha \; and \; \beta$ to be 1. The graph of hazard function was compared with the n-shaped hazard curve of B-S distribution, Birnbaum & Saunders (1969a).

The risk of becoming alcoholic was then evaluated using the equation 3.1

$$Risk \; of \; becoming \; alcoholic = 1 - \exp\left(-\int_0^t h(u)du\right) \qquad 3.1$$

Where, $h(u)$ is the hazard function, and *t* is the frequency of alcohol taking per week. The values of frequency of alcohol consumption per week in Table 4.2 were used as *t.*

The probability of developing AUD with time since onset was also evaluated. Results are displayed in Tables 4.5 and 4.6.

Non-parametric hazard rate was determined using logistic models and proportional hazards models. Hazard rate was estimated based on the measured variables themselves and also based on the factors formed from measured variables. In both cases event history method was used.

Data on duration since onset of alcohol taking to the date of interview was transformed from person level data to person period data for each respondent in order to apply discrete-time analysis.

3.3.1 Restructuring data for discrete-time analysis.

The data obtained using the questionnaire contained information on two times for each individual. The age at onset of alcohol consumption and the age when one was diagnosed as alcoholic or otherwise by interview date. This data was called person level dataset.

Before event history analysis was carried out the data had to be restructured so as to create a record for each episode. An episode is a continuous period during which an individual was at risk of experiencing an event, Steele (2005). Event history data arise, by following subjects over time and making notes about what happens and when. Usually this is done retrospectively and interest is concentrated to a few specific kinds of events.

The variable event is either zero or one. It is one when if the event (alcoholism) was observed on a respondent during the interview. The value zero indicated that the event was absent by the interview data and hence the respondent was right censored, (Brostrom, 2012). Each unit in an event history data set was presumed to enter the process at the same time. In terms of "calendar time," the time-of-origin may vary across observations, but in terms of "clock time," the starting point was generally treated as equivalent for all observations.

An assumption that there were no time varying covariates that is covariates remained unchanged throughout the observation period was made. A typical person level dataset that was made is shown below in Table 3.1. Ranking was done on the basis of increasing period, with the censored periods appearing last.

Table 3.1. Part of the person-level data set before transformation to a person-period data set

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Id	period	age_dig	diag years	years now	t diag	/ears onse	ageonset	Gender	agenow	race	relign	neighb
2	a104	35	5	35	48	0	35	31	1	2	1	1	1
3	a106	25	4	25	26	0	25	21	1	5	1	1	2
4	a133	24	3	24	34	0	24	20	0	3	1	1	2
5	a16	33	5	34	60	1	33	29	0	7	1	1	2
6	a17	34	5	34	60	1	33	29	0	7	1	1	2
7	a24	24	4	25	26	1	24	20	1	3	1	2	3
8	a24	25	4	25	26	1	24	20	1	3	1	2	3
9	a88	32	5	33	41	1	32	28	1	5	1	2	3
10	a88	33	5	33	41	1	32	28	1	5	1	2	3

3.3.2 Creating person period dataset

The first step in a discrete-time analysis was to create person-period data set. The event times and censoring indicator y_i , δ_i) were expanded to a sequence of binary responses $\{y_{ti}\}$. Where y_{ti} indicates whether an event has occurred in time interval $[t, t + 1)$. Thus for each observation, the event/censoring time y_i and the censoring indicator δ_i, and a time interval up to y_i, a binary response y_{ti}, was created, Steele & Washbrook (2013):

$$y_{ti} = \begin{cases} 0 & t < y_i \\ 0 & t = y_i, \delta_i = 1 \\ 1 & t = y_i, \delta_i = 0. \end{cases} \quad 3.2$$

Notably the respondents who did not have alcoholism by the interview date were censored while subscript i denoted the ith individual in this subsection.

Hence a sequence of binary responses was generated from each event time. Consequently, each individual's record was replicated as many times as the observed number of time intervals until either the event of interest or censoring occurred, (Stewart, 2010). Which lead to original dataset expanding in the size. These are shown in event column in Table 3.2. For example discrete responses are represented for respondent Id a225 are (0, 1). $Event = 0$ meant that the event was not experienced at age 31 while $event = 1$ meant that the event was experienced at age 32, one year after onset of alcohol taking. For respondent Id a24 whose discrete responses were (0, 0, 0) meant that from age of onset at 24 years and up to the time of interview at 26 years the respondent had not experienced alcoholism (become alcoholic). I.e. event was equal to zero for ages 24, 25 and 26 respectively, hence he was censored.

Table 3.2. Part of the person-period data set

	A	B	C	D	E	F	G	H	I	J	K	L
1	Id	Gender	PERIOD	age_dig	diag years	years now	D1	D2	D3	D4	D5	D6
2	a104	1	0	5	35	48	1	0	0	0	0	0
3	a106	1	0	4	25	26	1	0	0	0	0	0
4	a133	0	0	3	24	34	1	0	0	0	0	0
5	a16	0	1	5	34	60	1	0	0	0	0	0
6	a16	0	1	5	34	60	0	1	0	0	0	0
7	a24	1	1	4	25	26	1	0	0	0	0	0
8	a24	1	1	4	25	26	0	1	0	0	0	0
1444	a354	0	17	5	32	38	0	0	0	1	0	0
1445	a354	0	17	5	32	38	0	0	0	0	1	0
1446	a354	0	17	5	32	38	0	0	0	0	0	1

The column of period represented the duration between onset and either the diagnosis for those who had experienced alcoholism or the duration between onset of alcohol

taking and interview date for those respondents who had not experienced alcoholism by interview date.

For each respondent, the first row represented the age of his/her age of onset of alcohol taking. While the last row represented either the age they were diagnosed as alcoholic or the age at the time of the interview (due to right censoring), Danelia (2011). This type of censoring gave unbiased estimate of hazard rates as shown by Malacane, Murphy, and Collins (1997).

3.3.3 Estimating discrete-time hazards (life-table functions)

Discrete time hazards were calculated using equation 3.3, then plotted against the age at interview.

$$\hat{h}(t_{ij}) = \frac{number\ of\ events_j}{number\ \ at\ risk_j} \qquad 3.3$$

By treating discrete-time hazard as conditional probability, covariates and time were incorporated to form equation 3.4.

$$logitp_{ti} = \ log\left(\frac{p_{ti}}{1-p_{ti}}\right) = \boldsymbol{\alpha}(t) + \boldsymbol{\beta}'\boldsymbol{x_{ti}} \qquad 3.4$$

Where $\boldsymbol{x_{ti}}$ is a vector of covariates and $\alpha(t)$ is a function of t.

This was done by creating a rectangular dataset containing respondent's time and covariates by introducing k-1 temporal dummy variables. Model fitting was done using Rcmdr package, Version 2.5-3 in R. Dependent variable was the dichotomous event indicating whether or not the respondent had AUD at the time of interview. Logistic regression, using generalized linear model (glm) with the family being binomial and link function, logit was applied.

The first model ("baseline model") consisted only the time period with no intercept. Several other baseline models were investigated. A full model was formed consisting of all the variables applied at once. Then, stepwise model selection was applied opting for backward shift method.

The second approach involved combining factors (generated from measured variables) to form equation 3.5.

$$Y = \beta_1(age\ of\ onset) + \beta_2(personality) + \beta_3(family\ attention) + \beta_4(environmental) + \beta_5(social - cultural) + \beta_6 ES + \beta_7(drinking\ habit\ and\ patterns) + \beta_8(gender) + \beta_9(peer\ pressure) \quad 3.5$$

Where Y is the dependent variable, alcoholism and $\beta_1, \beta_2, \beta_3, \ldots, \beta_9$ are coefficients.

The coefficients $\beta_1, \beta_2, \beta_3, \ldots, \beta_9$ were estimated using the Maximum Likelihood Method. They were then used to construct an estimate of the baseline hazard function by substituting equation 3.5 in equation 3.4.

Logistic regression was elected for modelling the outcome, using generalized linear model (glm) with family being binomial while link function was logit. Then stepwise model selection was applied using backward shift method. Effect of gender in particular was later analysed in greater details.

3.4 Projecting Incidences of Alcoholism

To reconstruct the incidences of alcoholism linear logistic function was used while back-calculation was used to evaluate the incidences of alcohol use initiation. Projections were made using back-calculation method. MLE was used to estimate parameters in the logistic model.

3.4.1 The logistic model

Having assumed a strong parametric form for $f(t;\alpha,\beta)$ in equation 3.11, the same was done for incidences of alcoholism, $a(t)$, where appropriate approximation to the stochastic process that gave rise to alcoholic incidences was assumed to be Poisson process model. With the respective incidences being events in assumed Poisson process. As in Simwa and Pokhariyal (2003) the incidence model was assumed to be linear logistic function given as,

$$a(t;a_0,a_1,a_2,a_3)=\frac{a_0+a_1t}{1+e^{(a_2-a_3t\hat{}2)}} \qquad 3.6$$

MLE method was applied to estimate parameters a_0,a_1,a_2,a_3. And Pearson Statistic was used to measure goodness of fit for the incidence model 3.6.

For annual projections time t was represented by the set $T=\{1,2,3,4,5,6\}$. While for quarterly projections time t was represented by the set $T=\{1,2,3,\dots,23\}$, where $T\subset U$ is the interval when data is not confounded. With $U=\{1,2,3,\dots,n\}$.

To project the number of alcoholic incidences annually, values of t were taken as $t=\{7,8,9\}$ while t was taken as $t=\{24,25,26\}$ for quarterly cases.

It was assumed that alcoholism incidence data contained information about the most recent exposures. Therefore numbers of individuals abusing alcohol in the last few years were reliably reflected in the alcoholism incidence data.

3.4.2 The Back-calculation method

The exposure times were modelled using the function $\lambda_s(t,\boldsymbol{\beta})$ where, t is time-dependent variable and $\boldsymbol{\beta}$ a vector of parameters, Egan & Hall (2015).

While the pdf of times of diagnoses was given as

$$a(t;\boldsymbol{\alpha},\boldsymbol{\beta}) = \int_{T_0}^{T_L} f(t-s;\boldsymbol{\alpha})\lambda_s(t;\boldsymbol{\beta})\, ds \qquad 3.7$$

Where, $f(t-s;\boldsymbol{\alpha})$ is the incubation period distribution.

Equation 3.7 was expressed in matrix form as follows:

Let A_i *be the number of alcoholics dignosed* in the interval $(t_{i-1}, t_i]$ and $f(j,i)$ be the probability of becoming alcoholic in the interval $(t_{j-1}, t_j]$. Assume this probability remains stationary.

Define a non-empty set, $U = \{1,2,\dots,j\}$.

As in Becker (1990), relation between new yearly number of exposures to alcohol use and the annual number of alcoholics diagnosed is given as

$$E\left[A_i/H_1, H_2, \dots, A_j\right] = \sum_{i=1}^{j} H_i f_{j-i}\, for\ all\ i\ \in U \qquad 3.8$$

Where H_i is the number of newly initiated alcohol users in the time interval $(t\{j_1\}, t_j]$ and

$H_1, H_2, \dots H_n$ is assumed to be a typical discrete non-homogenous process. Taking expectations on both sides gives equation 3.9.

$$a_i = \sum_{j=1}^{i} \lambda_i f(j,i) \quad for\ all\ i\ \in U \qquad 3.9$$

Where $\lambda_i = E[H_i]$ and $a_i = E(A_i)$.

Equation 3.9 is the discrete form of equation 3.7 obtained by replacing probability densities for number of diagnosed alcoholics and expected number of exposures, respectively.

λ_i was solved using the matrix method in equation 3.10 by inverting matrix $[(f(j,i))]$

$$[(a_i)] = [(f(j,i))]\lfloor(\lambda_i)\rfloor \qquad 3.10$$

The order of $[(a_i)]$ and $\lfloor(\lambda_i)\rfloor$ $is\ n \times 1$ while $[(f(j,i))]$ $is\ an\ n \times n\ square\ matrix.$

$[(f(j,i))]$ is the probability matrix obtained by integrating $f(t;\alpha,\beta)$, the incubation distribution

$$f(t;\alpha,\beta) = \frac{1}{2\sqrt{2\pi\alpha\beta}}\left[\left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right] exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right],$$

$$0 < t < \infty\ ;\ \alpha, \beta > 0 \qquad 3.11$$

$$f(j,i) = \int_0^{t_i-,t_j} f(t;\alpha,\beta)\, dt \qquad 3.12$$

One key assumption is that this distribution which was determined using data from seven counties was applicable for data relating to the whole country, Kenya. Hence it was applied in the back-calculation model to project alcoholism cases for the entire country.

Matrix $[(a_i)]$ was a row matrix formed using the data obtained from Ministry of Health relating to morbidity (diagnoses with alcoholism) from 2014 to Sept 2019.

The estimates of λ_i represent the number of individuals who initiated alcohol taking since the start of epidemic (an arbitrary time taken before first diagnoses of an alcoholic).

The projected number of alcoholics during the year $i + \tau$ is given by

$$a_{j+\tau} = \sum_{i=1}^{j+\tau} \lambda_i f_{j+\tau-i} \qquad 3.13$$

CHAPTER 4: RESULTS AND DISCUSSIONS

4.0 Introduction

In this chapter the results of structural analysis of the relation between alcoholism and its risk factors, characteristics of the model of incubation period alcoholism, parametric and non-parametric hazard rates, probability of becoming alcoholic, hazard and survival probabilities, logit-based models for hazard function, effect of gender on hazards and prediction of alcoholism incidences are presented and discussed.

4.0.1 Demographic characteristics of the sample

Of the 780 study participants (594(76.2%) males and 186 (23.8%) females), 210 were from rural areas, 520 from urban centres and 50 from rehabilitation centres. Twenty five (3%) respondents were excluded from analysis for the following reasons. Eighteen (2%) non-response cases which arose from refusal to respond, three gave information that contained contradictions (0.5%). Four (0.5%) had crucial information missing due to refusal to answer crucial areas and incomprehensiveness of their response.

Thus 755 respondents (96.8% completion rate) provided valid data. Of these 755 respondents males were 579 (76.3%) and females 176 (23.7%). The age of respondents by the interview date ranged from 14 years to 65 years with median age being 32 years, while age at diagnosis ranged from 14years to 52years. Out of 755 respondents 235 (31.1%) met criteria/conditions for classification as having AUD, while 520 (68.9%) had not by date of interview.

4.1 The structural analysis of alcoholism and its risk factors

The basic characteristics the primary data such as the means, range and median were explored and are presented as follows.

4.1.1 Exploratory data analysis of independent variables.

Characteristics of ages are given by distributions for the age at onset shown in figure 4.1 and age at interview in figure 4.2.

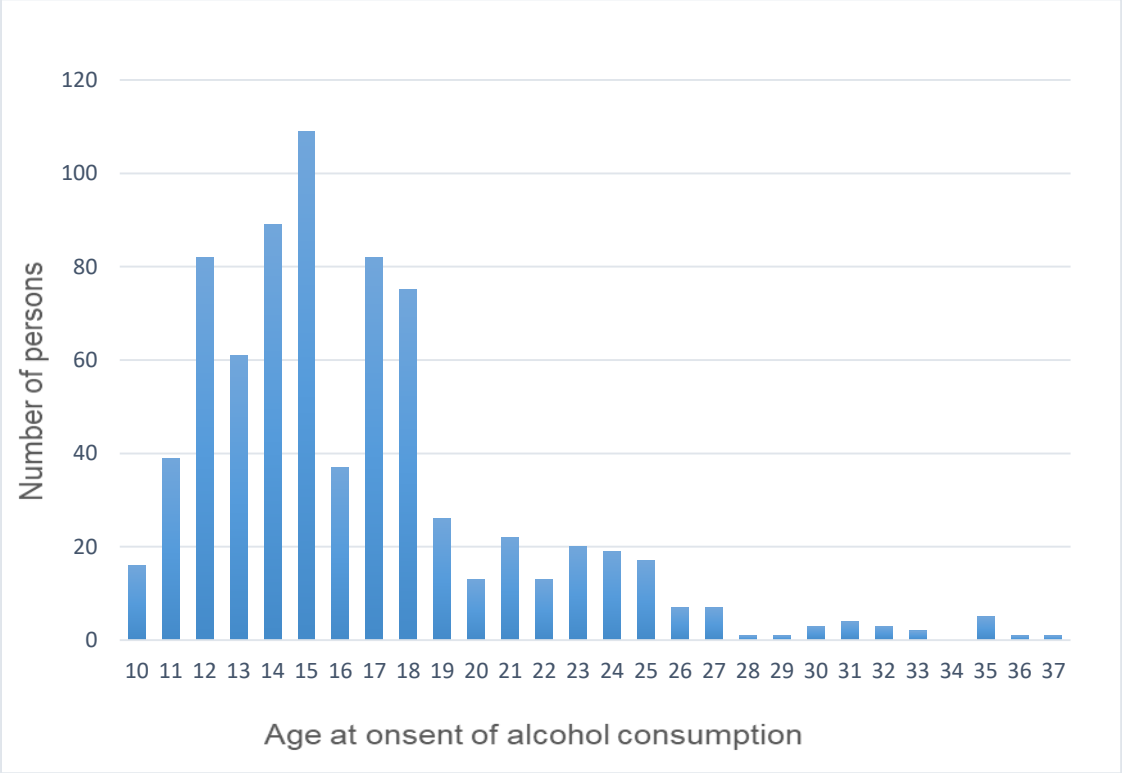


Figure 4.1 Graph of age at onset of alcohol consumption

The mean age at onset of alcohol taking was 16.53 years and median was 15 years. The earliest age at onset was 10 years and oldest was 37 years.

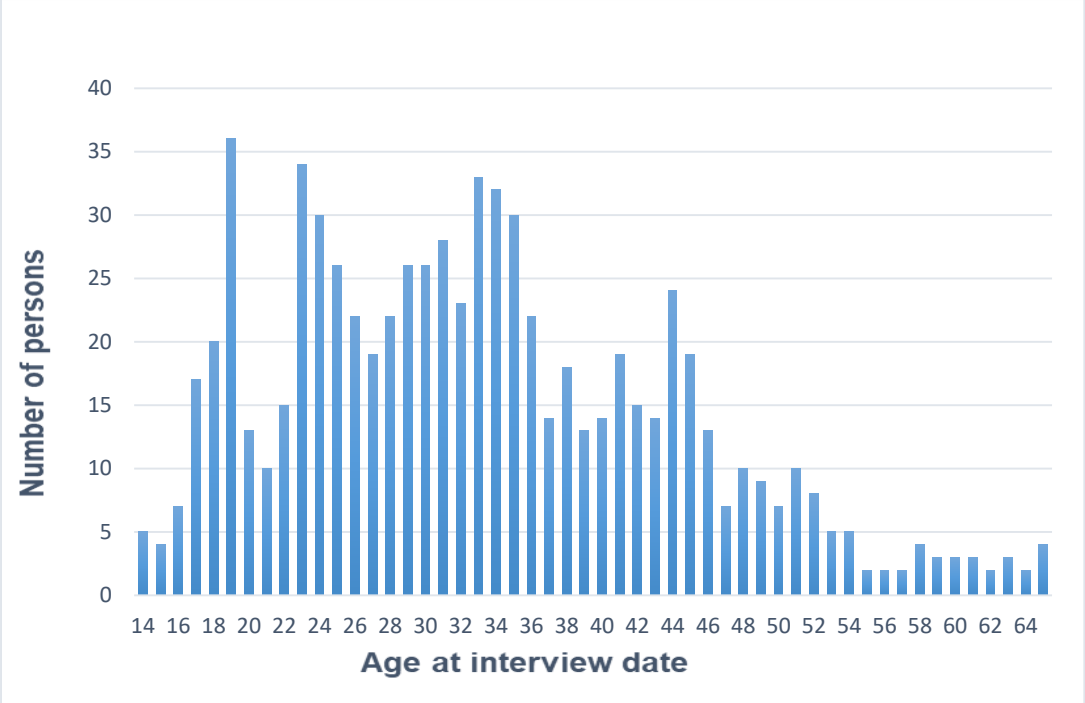


Figure 4.2 Graph of age at interview

The mean age at interview which was 33.25 and median age was 32 years. The oldest interviewee was 65 years old while the youngest was 14 years.

The correlation between the items is presented in a Correlogram in figure 4.3.

The factors in figure 4.3 are arranged such that rows and columns have been reordered using principal components analysis to cluster variables that have similar correlation patterns together. The darker the colour the greater is the magnitude of correlation.

Blue colour shows a positive correlation between the two variables that meet at the cell. Whereas, Red colour shows a negative correlation between the two variables that meet at the cell.

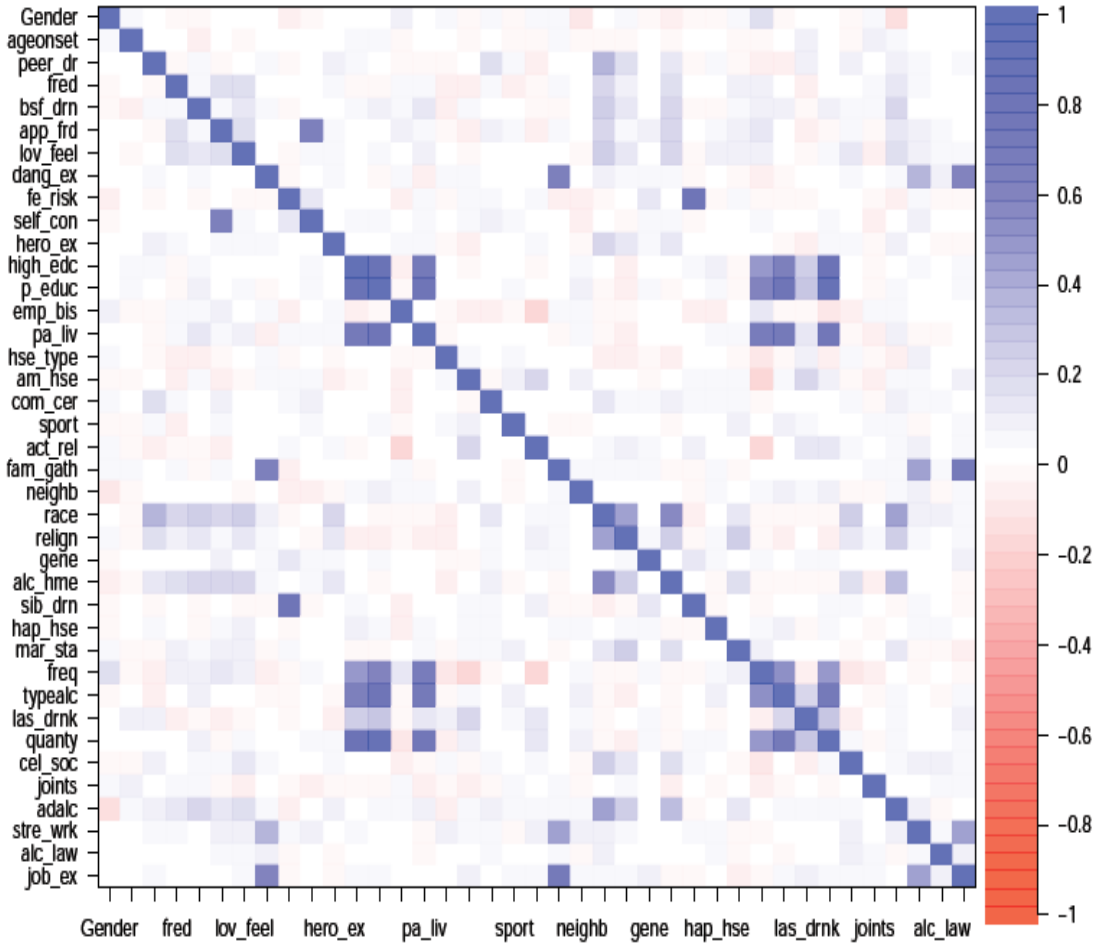


Figure 4.3 Correlogram illustrating diagrammatically the correlation between measurable variables

Note. The key for coding on the variables is in Appendix E2

4.1.2 Suitable number of factors and the Model

Parallel analysis suggests that the number of factors be 13 and components be 11. The se 11 components contain 54.6% of variance. 13 factors have chi square value of 321. 9 on 312 degrees of freedom. This is illustrated in the scree plot in Figure 4.4.

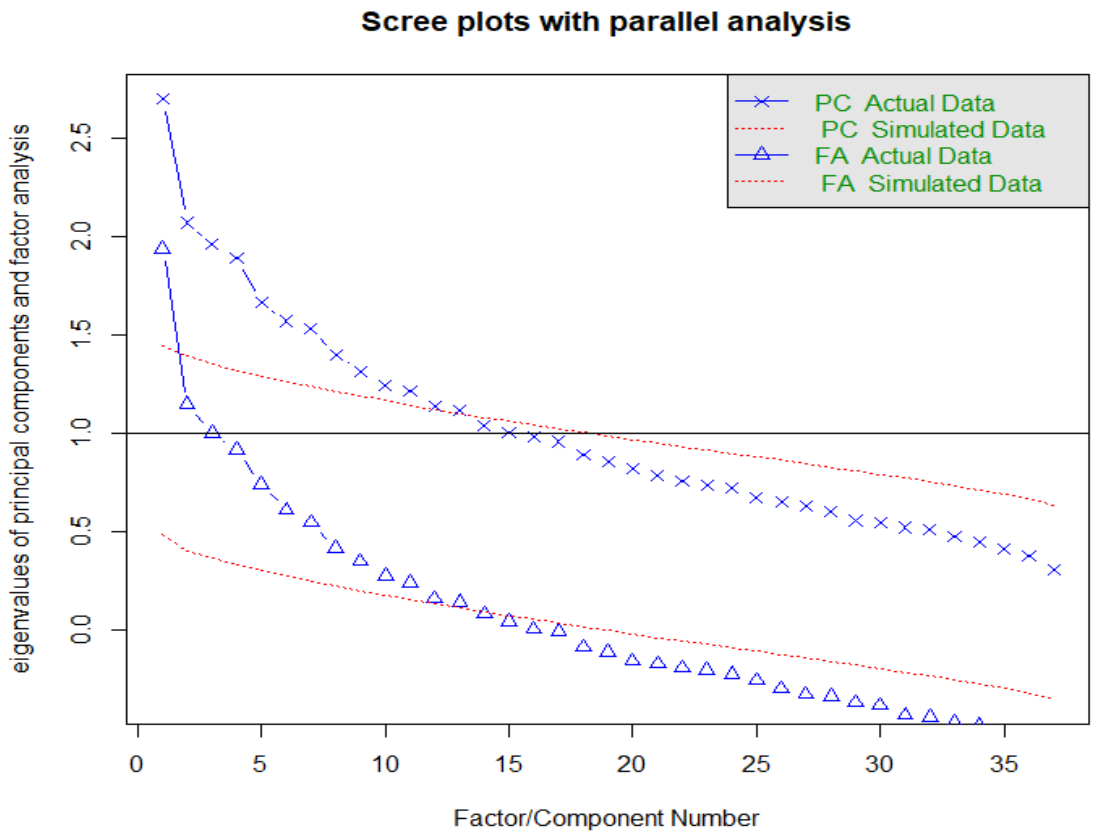


Figure 4.4 Scree plot for both actual data and simulated data

Using the protocol suggested by Brown (2006) and Schreiber, et al., (2006) for reporting Structural Equation Modelling (SEM) analyses, the results were summarized in table 4.1

a. RMSEA (Root Mean Square Error of Approximation) which measures fit index for adjusting for model parsimony. The cut-off value is 0.06. Those close or below it are taken to give a good fit. Values for RMSEA were all within this, showing that the conceptualized model fits reasonably well with our population.
b. CFI, the cut-off is 0.9.values above this indicate high average correlation between variables. Our results for CFI were at 0.8 for all models. This indicates average correlation is not very high.

Table 4.1: Summary results of Model Fit indices

Indices	Conceptual model	Intermediate model	Improved model	Final model	CUT-OFF
TLI	0.745	0.728	0.732	**0.776**	> 0.95
CFI	0.774	0.758	0.766	**.803**	> 0.95
IFI	0.776	0.761	0.769	**0.805**	
RMSEA	0.061	0.062	0.061	**0.060**	< 0.07
SRMR	0.060	0.058	0.058	**0.059**	< 0.08
ECVI	3.073	2.995	2.941	**2.552**	
AIC	39826.680	39294.274	39414.446	**35102.42**	
X^2 *(df)*	2104.1 (558)	2057.2(528)	1998.2(595)	**1733.4(464)**	

c. SRMR (Standardized Root Mean Square Residual). Measures absolute fit of the data in relation to theoretical model. Values close to or below 0.08 indicate good fit. Our values were 0.06 for all the models. Thus data fitted well in relation to the theoretical model.

d. TLI (Tucker Lewis Index) describes the comparative fit of a data set with theoretical model. Values above 0.9 indicate a good fit. Our models produced values between 0.7 and 0.8. Thus the fit was not very strong.

4.1.3 The final SEM model

The Structural Equation Modelling produced figure 4.5 as the most suitable linkage be tween factors and related latent variables.

Peer influence was linked to peers drinking, friends drinking when together, best frien d drinks and one approved it. Drinking habits and patterns are linked to how often one take alcoholic beverage, type of alcoholic drink normally taken, how many drinks one normally take in a single occasion and when one last took alcoholic drink.

Economic status linked to interviewee's education level, what parents/ guardians do fo r a living, parent's or guardian's highest education level, being employed or in busines s, house having basic amenities such as clean piped water and electricity, living in a p ermanent house or house type.

Factors describing personality includes love feeling of being drunk, drinking promotes self-confidence, drinking excess alcohol shows you are a hero, and I ignore warning a bout dangers of excessive drinking. Family and family attention was linked to my pare nts/guardian/siblings drink alcohol, marital status and happiness, peace, fun or joy at h ome. Socio-cultural was linked to involvement in sports, race, involvement in religiou s activities and community having occasions/ceremonies where alcohol taking by all is permitted.

The path model shown by Figure 4.5 depicts the causal relations between risk factors of alcoholism (variables). This model was read from top to bottom, with the variables on the top (independent variables) predicting the outcome variable on the bottom. The path coefficient indicated the direct effect of a variable assumed to be a cause on another variable. These coefficients are standardized because they were estimated from correlations. Some coefficients don't meet the threshold set by Huber et al. (2007), "Path coefficient must be at least 0.100 and at a significance level of at least 0.05"

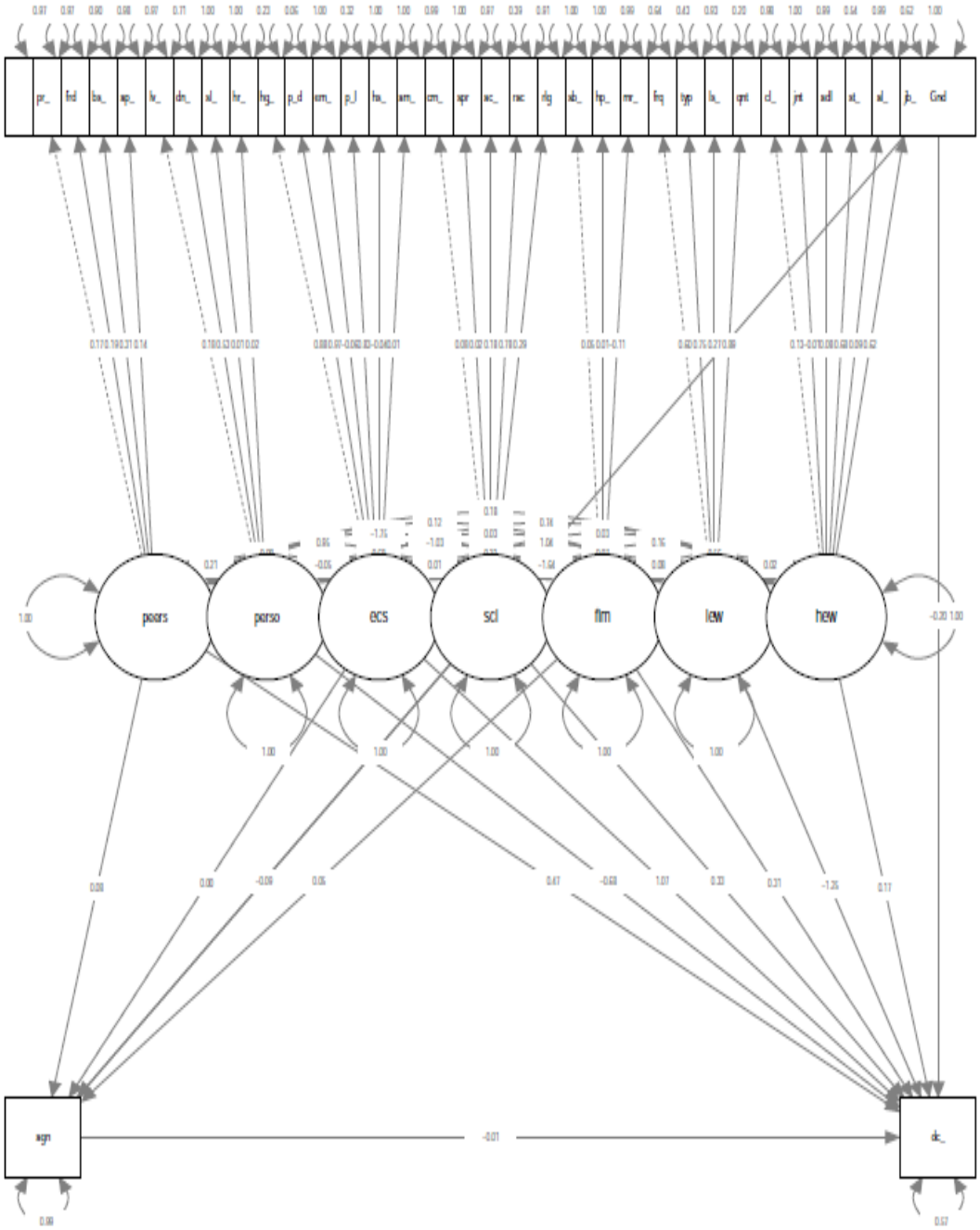


Figure 4.5: Final Path model showing factor loadings and relationships between variables

Comparing the conceptual path diagram, Figure 3.1 to the final path diagram Figure 4.5, the following observations were made. First was the creation of causal relationship between factors. These include gender had an effect on drinking habits and patterns. Job or occupation or daily activity influence age at onset of alcohol taking. Peer influence, economic status, family and familiar attention influenced the age at onset of

alcohol taking. Drinking habit and patterns was influenced by eight factors. Secondly two more factors (latent variables) were created; lew and hew. Lew comprised of freq, quanty, typealc and lov_feel while hew consisted of dang_ex, joints, adalc, job_ex, act_rel, and alc_hme. Thirdly, the factor lew and drinking habits and patterns influence each other (covariance = -1.25).

4.2 Model for Incubation Period of Alcoholism

The results are in two parts. In part one the theoretical bases or foundation for selecting B-S distribution in modelling incubation period of alcoholism are presented. Finally in the empirical results where the model is fitted to the frequency of alcohol taking data were presented in part two.

4.2.1 Theoretical basis for selecting B-S distribution

The following theoretical reasons for selecting B-S distribution to model incubation period of alcoholism were gathered from literature.

a) There is evidence of cumulative damage from stress caused by excessive alcohol consumption as illustrated in the introduction and in sections 2.2.1 and 2.2.2. From the results of various researches described in the above sections it is clear that biophysical processes leading to alcoholism support the use of B-S distribution in modelling incubation period of alcoholism. In particular, the products of alcohol metabolism (acetaldehyde and acetate) are very reactive. The formation of acetaldehyde induces toxic effects by binding to protein and DNA resulting in functional alterations. Besides alcohol significantly alters gut micro-biota resulting in an altered balance of pathogenic and commensal organism, ultimately leading to inflammatory process.

The expected fluctuation in BAC of an individual who consumes alcohol on a regular basis results in stress and tension in the cells and organs causing structural damage to occur. This is illustrated theoretically in Figure 4.6. Most systems operate within a given tolerance limit before failure. Assume the cell or system or organ started deteriorating at point A but the person may not seek physician's help until B. During this duration (A-B) a person may be asymptomatic. X denotes the calendar time of onset of drinking, Z denotes the chronological time when an individual gets diagnosed as alcoholic. The duration between X and Z is the incubation period.

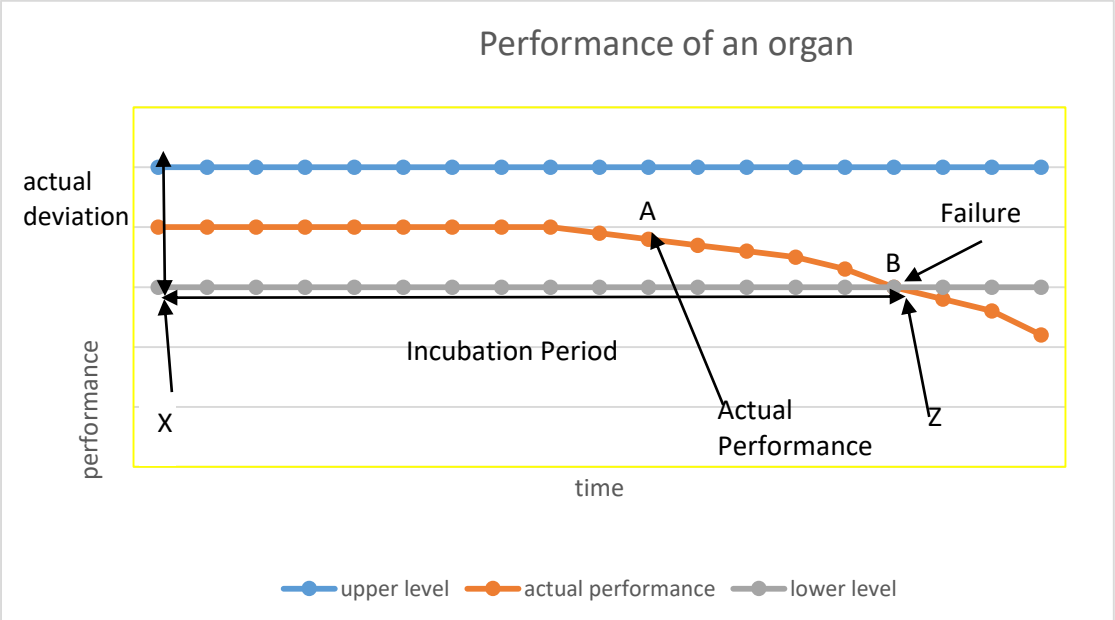


Figure 4.6 Theoretical representation of performance of an organ or cell with time

Cumulatively these damages affect performance of the concerned cells and organs resulting in irreversible harm and death of cells. According to Leiva et al.(2008)This, argument supporting "cumulative damage" or "fatigue" is a strong reason to invoke use of Birnbaum-Saunders distribution in describing lifetime data,.

b) Consider an individual who habitually consumes alcohol on a regular basis. Each time he/she takes alcohol his/her BAC rises and falls causing cells/organ involved in alcohol metabolism to be exposed to a sequence of stress. This is typical of persons who have developed tolerance to alcohol. BAC curve (absorption and elimination of alcohol) as discussed in section 2.1.1, may form pattern shown in Figure 4.7 in the long run. Let λ be the mean BAC. Assume that alcohol levels below λ have no negative or damaging effect to the body. As seen earlier, that negative effects of alcohol consumption occur when one exceeds a certain threshold.

Assume that each drinking session last any length of time irrespective of the quantity of a drink provided there is no break of more than four hours.

Let BAC level at any time interval be indicative of the amount of stress (load) imposed on the cell. Thus at the imposition of each load, damage is extended by a random amount. Cell damage at each cycle is unobservable in practice we get to know of it when organ failure occurs (either patient visits the doctor or dies or both). It was assumed that the following conditions apply.

- The cell is subjected to a sequence of cyclic loads (rise and fall in BAC during alcohol absorption and elimination process) which produce / cause damage.
- Failure occur when the amount of cell damage (area or volume) exceeds certain level (ω) or when its functionality is altered. Or the number of damaged (dead) cells in an organ exceeds a certain number/ level/quantity.
- The total size of damage due to the jth drinking session Y_j is a random variable that follows a statistical distribution with mean, μ and variance, σ^2.

B-S model seeks the distribution of the smallest number of sessions, $say\ n^*$ such that:

$$S_n = \sum_{j=1}^{n} Y_j \qquad 4.1$$

$of\ \ n\ positive\ random\ variables,$ exceed the given threshold ω i.e.

$$n^* = \inf\left\{ n \in N: \quad S_n = \sum_{j=1}^{n} Y_j \ > \omega \right\} \qquad 4.2$$

Let N be the number of required drinking sessions until failure.

Birnbaum and Saunders (1969) showed that

$$P(N \leq n) = \Phi\left(\frac{\sqrt{\omega\mu}}{\sigma} \left[\sqrt{\frac{n}{\frac{\omega}{\mu}}} - \sqrt{\frac{\frac{\omega}{\mu}}{n}} \right] \right) \qquad 4.3$$

This scenario is similar to the one envisaged by Desmond (1985) in his argument that if the response variable (in this case BAC) is above a fixed level for too long, then time to failure belongs to a B-S distribution family.

c) The theoretical arguments applied in establishing the genesis of the BS distribution mirror the process alcohol uptake. Fatigue life for a cell or organ is the lifetime equivalent to the number of cycles (above average drinking sessions) until the failure due to fatigue.

Let the rise and fall in BAC be visualised as a continuous unimodal function defined on a unit interval. Fatigue is weakening of the cellular material resulting in structural damage which is accumulated by the organ. Failure due to fatigue occur when the cumulative damage exceeds a threshold level of functionality of the specimen. Organ

failure is caused by ultimate death/deformation or depletion of cells that make up this organ. Assume that the rate of cell damage or death is higher than the rate of cell repair. Assume also that cell fatigue process which leads to organ failure, occur through the following three stages:

Fist an imperceptible damage in the cell begins to form followed by the growth and propagation of damage in the cell, which creates an irreversible state in it or material or its function due to cyclic stress (caused by chronic alcohol use or resulting in AUD) and finally the failure of the material/ cell death due to fatigue occurs.

Let Figure 4.7 below represent typical variation of BAC with time for a regular social alcohol consumer.

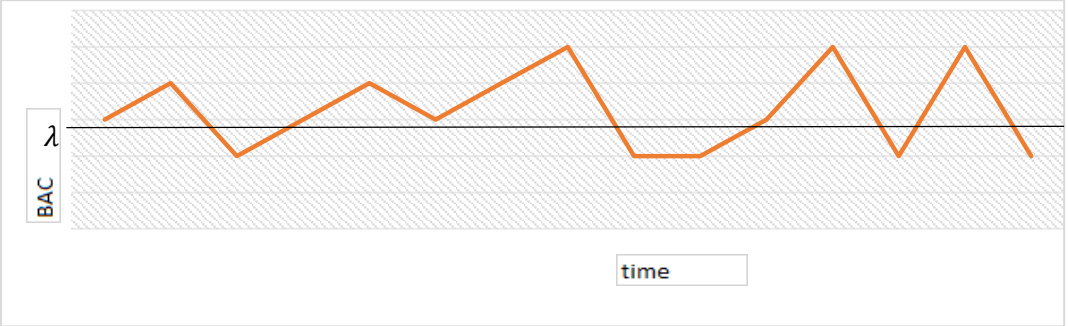


Figure 4.7 Variation of BAC with time as it crosses the mean position (threshold for damage)

As in Desmond (1985), let

$$D(t) = \begin{cases} \sum_{i=1}^{n} T_i & U(t) > 0 \\ 0 & U(t) = 0 \end{cases} \qquad 4.4$$

Where $U(t)$ is the number of times the BAC curve rises above the mean level λ.

Note T_i is a factor of peak BAC. Therefore it is a factor of number of drinks and time which is actually the frequency or rate of drinking (assuming concentration of alcohol in a drink remains constant during all drinking sessions).

Where $U(t)$ is the number of crossings above the λ level,

$U(t) = 0, 1, 2, \dots, n\ in\ interval\ (0, t)$.

Further, as in Desmond (1985), let T denote time to becoming alcoholic, with $T > t$ $if\ and\ only\ if\ \ D(t) < C$, then the probability density function of T is

$$f(t) = \frac{1}{\sqrt{2\pi}}\left(\frac{C}{2B}t^{-3/2} + \frac{A}{2B}t^{-1/2}\right) exp\left\{-\frac{1}{2}\left(\frac{C - At}{B\sqrt{t}}\right)^2\right\} \qquad 4.5$$

And cumulative distribution function is

$$F(t) = \Phi\left(\frac{At - C}{B\sqrt{t}}\right) \qquad 4.6$$

Where:

A is determined by the mean of $D(t)$

B is determined using the variance of $D(t)$

C is a fixed amount for which BAC is greater than λ

Which belongs to the B-S family of distributions, Desmond, (1985).

The theoretical arguments applied in establishing the genesis of the BS distribution mirror the general uptake of alcohol and thus it is an appropriate model for describing alcoholism phenomena. In addition the cell damage arising from stress due to excessive

alcohol use satisfies conditions in Desmond (1985), confirming the suitability of use of B-S distribution in modelling time from onset of alcohol taking to becoming alcoholic.

4.2.2 Maximum Likelihood Estimate of $\alpha \, and \, \beta$

Several methods of estimation of $\alpha \, and \, \beta$ are available. Birnbaum and Saunders (1969) obtained the MLEs of $\alpha \, and \, \beta$, then proposed the initial estimate of β to be the mean-mean estimate. Dupuis and Mills (1998) gives other estimation methods of $\alpha \, and \, \beta$.

The MLE of the parameters α, β in the function

$$\mathrm{f}(t;\alpha,\beta) = \frac{1}{2\sqrt{2\pi\alpha\beta}}\left[\left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right] exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right] \;\; , \; 0 < t < \infty \; ; \; \alpha, \beta > 0 \quad 4.7$$

are obtained by solving the non-linear equation for $\hat{\beta}$

(Ng, Kundu, & Balakrishnan, 2003)

4.2.3 Empirical analysis of the model

First the descriptive characteristics of the data are presented, then followed by estimates of the model parameters and graphs of Probability density functions (pdf) and Cumulative density functions (cdf), and finally measures of goodness of fit indices were given.

4.2.3.1 Descriptive summary of the respondents

Data on frequency of alcohol consumption per week was summarised in table 4.2. A session was defined as a non-interrupted period of alcohol taking. Where a break of more than four hours exists between drinks, then a new session arises.

Table 4.2. Summary of frequency of alcohol taking the week preceding interview

Actual drinking frequency per week	1	2	3	4	5	6	7	8	9	10
Number of persons	7	17	26	15	15	12	28	44	25	7

Actual drinking frequency per week	11	12	13	14	15	16	17	18	19	20	21
Number of persons	10	7	5	4	2	2	2	4	1	1	1

The amount of alcohol taken and length of a session were not taken into account. Thus one who drank continuously for say half an hour and one who drunk continuously for say five hours were considered to have one session regardless of the amount drunk. Figure 4.8 was the graph drawn to represent the frequency of drinking per week.

The data on frequency of alcohol taking per week had the following characteristics.

235 (31.1%) respondents were identified as suffering from alcoholism. Their mean frequency of drinking per week was 7.99 times. The median was 8, mode of 2 and 10, minimum and maximum of 1 and 21 respectively. Skewness of 0.59, kurtosis of 0 and standard error of 0.3

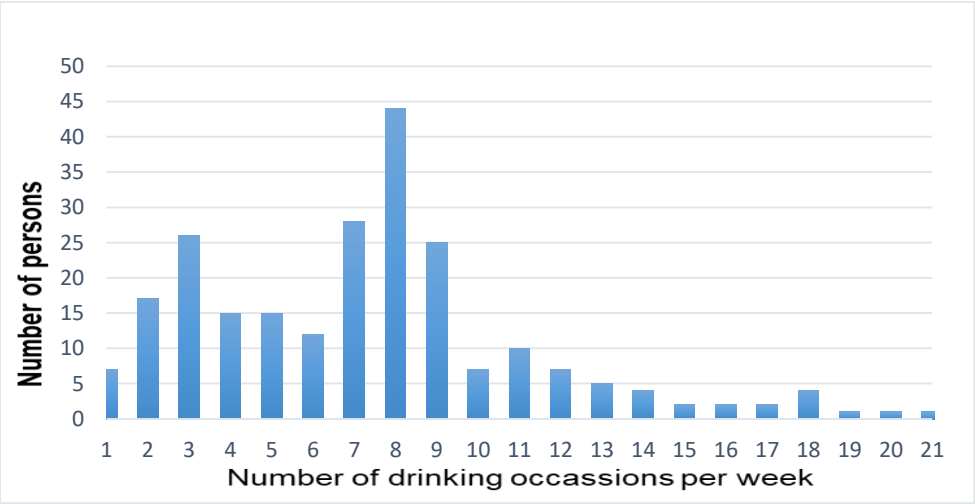


Figure 4.8: Graph of frequency of drinking per week

4.2.3.2 The incubation period distribution and estimates of α and β

The MLE's are: α = 0.76584 and β = 6.1328. Their 95% confidence intervals are:

Table 4.3: Parameter estimates with 95% confidence intervals

Parameter Estimates	Value	Lower limit	Upper limit
α	**0.7658438**	0.68	0.85
β	**6.132811**	5.44	6.83

These values of α and β were fitted to incubation period distribution which is the pdf of B-S distribution given by equation (4.8). Cdf is given by equation (4.9).

$$f(t;\alpha,\beta) = \frac{1}{2\sqrt{2\pi\alpha\beta}}\left[\left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}}\right] exp\left[-\frac{1}{2\alpha^2}\left(\frac{t}{\beta} + \frac{\beta}{t} - 2\right)\right] \quad ,$$

$$0 < t < \infty\ ;\ \alpha,\beta > 0 \qquad 4.8$$

$$F(t;\alpha,\beta) = \Phi\left(\frac{1}{\alpha}\xi\left(\frac{t}{\beta}\right)\right) \quad t > 0, \alpha > 0, \beta > 0 \qquad 4.9$$

Estimates of α and β obtained using mlebs(x), mmebs(x) and gmebs(x) functions in R Studio, Version 1.2.1335 are given in Table 4.3. Estimates of alpha and beta are similar under MLE and MME methods but quite different under graphical method.

Table 4.4. Different estimates of α and β

Method	Estimate of α	**Estimate of β**
Maximum Likelihood Estimate	0.7658438	6.132811
Modified Moment Estimate	0.7657952	6.179523
Graphical method	0.6021526	6.697022

4.2.3.3 Pdf, cdf of B-S distribution

The graphs of probability density function (pdf) and cumulative distribution function (cdf) of B-S distribution with different values of α *and* β are represented in Figure 4.9 and Figure 4.10 respectively. The shape of the curves are typical of B-S distribution curves. Results of fitting frequency of alcohol taking per week in the B-S distribution confirms that Birnbaum-Saunders distribution is well suited to model time from onset of alcohol taking to diagnosis with alcoholism.

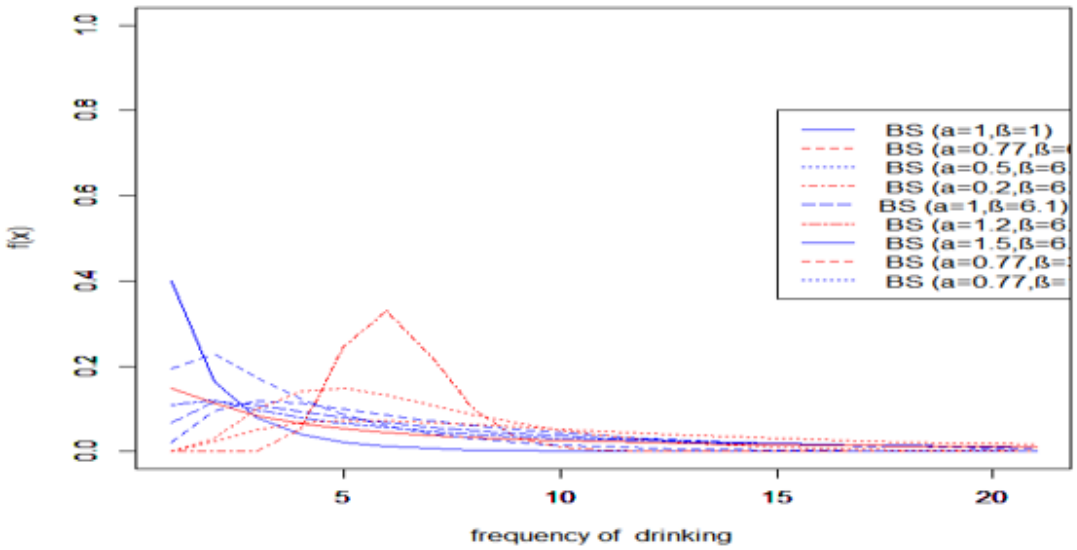


Figure. 4.9 Probability density function of B-S distribution

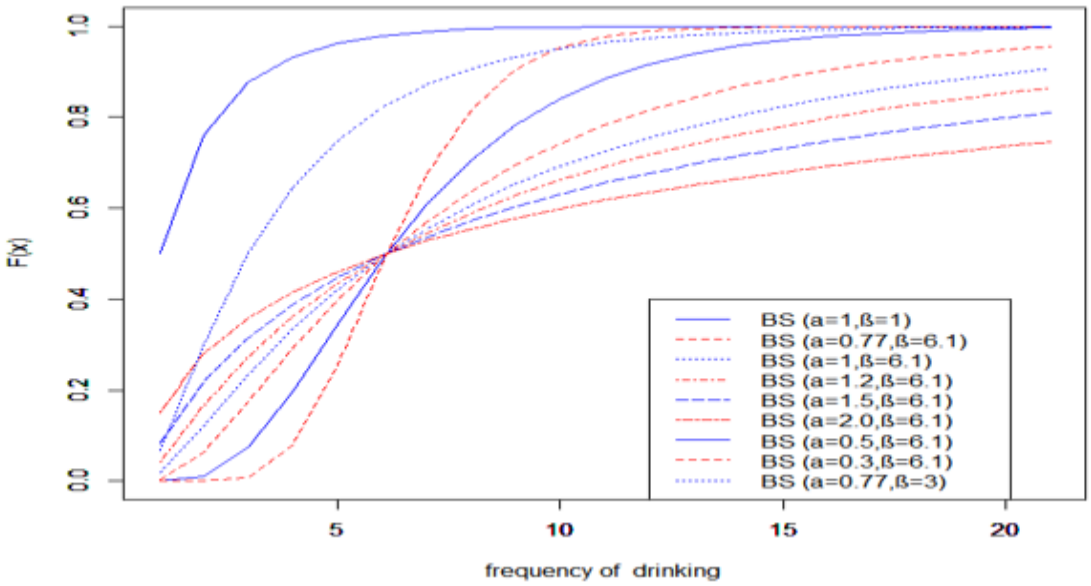


Fig. 4.10 Cdf plot for the B-S distribution based on frequency of alcohol intake per week.

The B-S distribution has an upside down curve (n-shaped) for all values of alpha and beta.

4.2.3.4 Goodness of fit.

The results of pp- plot, QQ- plot, TTT plot, r-squared, K-S test, AIC, are presented.

The pp-plot

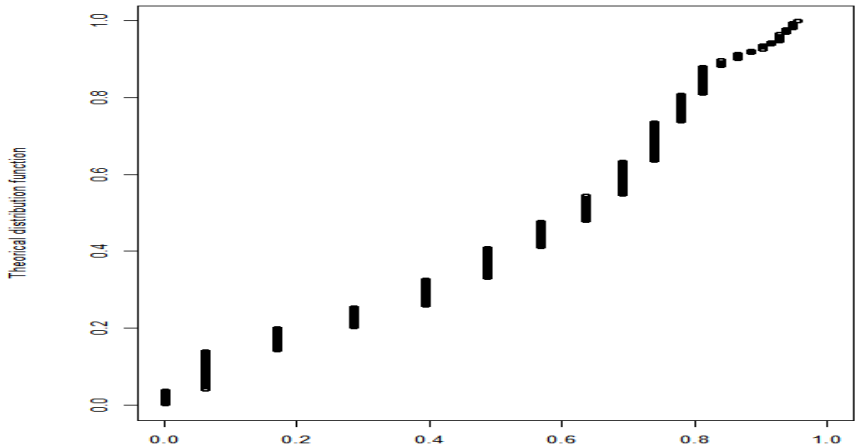


Figure 4.11. The p-p plot based on standardised values

The P-P plot compared the empirical cumulative distribution function of a frequency of alcohol taking with a theoretical B-S cumulative distribution function. Since the overall pattern of Figure 4.11 follows approximately a straight line, then the data followed the assumed (B-S) probability distribution.

The Q-Q plot

The Q-Q plot, Figure 4.12 compared the quantiles of data of frequency of drinking per week distribution with the quantiles of a standardized theoretical distribution from a B-S family of distributions. The purpose of Q Q plots was to find out if two sets of data come from the same distribution. A 45 degree angle is plotted on the Q Q plot, Figure 4.13; if the two data sets came from a common distribution, the points would fall on that reference line.

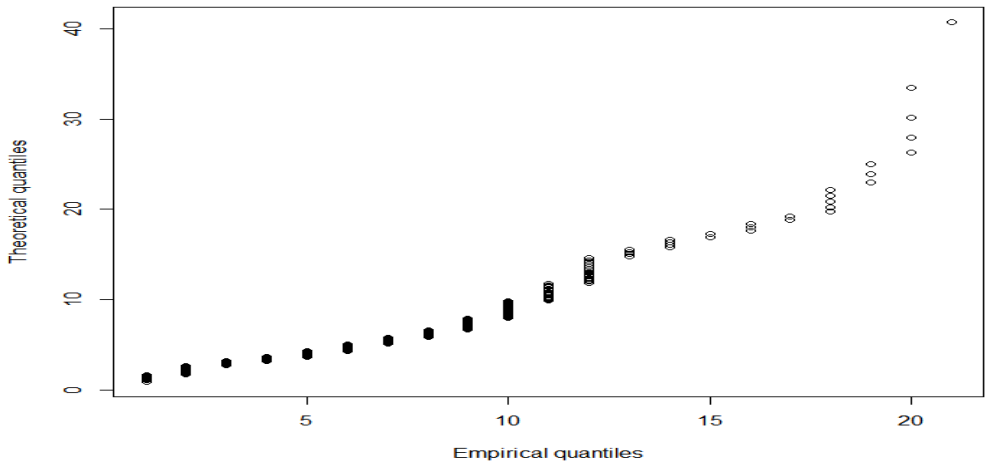


Figure 4.12. The q-q plot

Fitting the data graphically generates the values of alpha and beta as follows

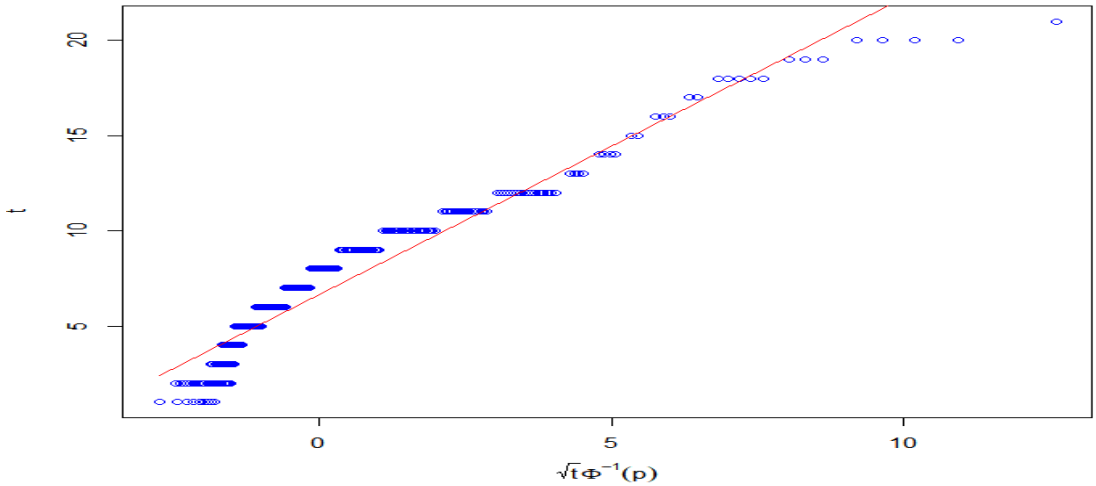


Figure 4.13: Graphical method of testing goodness of fit

Figure 4.13 show Q Q plot (also called normal quantile-quantile (QQ) plot) with a 45 degree line. The points are not clustered on the 45 degree line at the tails, suggesting that the sample data is not normally distributed..

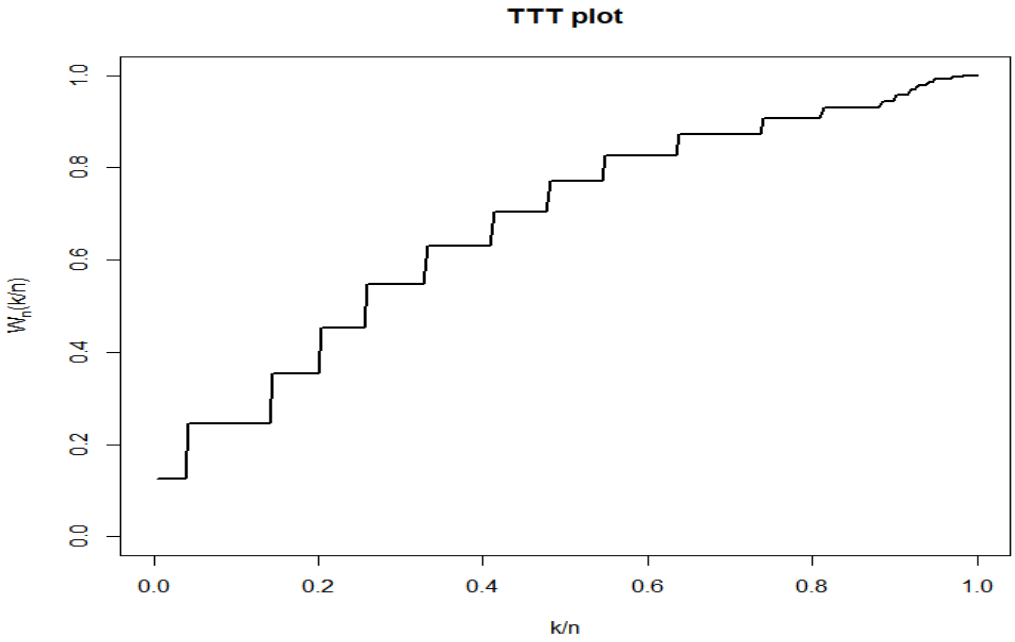


Figure 4.14 The TTT plot of the BS distribution.

The Total Time on Test Plot (TTT-plot) Figure 4.14, was introduced as a tool for a test quantity for testing exponentiality of failure data. For perfectly exponential curve data

the plot would be passing through the diagonal drawn from the origin. This plot was slightly above indicating that data is not from exponentially distributed population.

The results of other tests of fit for were given below:

1. The one sample Kolmogorov-Smirnov test gave a value of D=0.16094 and a c orresponding p-value of p = 0.00001033. This shows that the data fits the assu med distribution significantly well.
2. R^2 = 0.9356703, thus there is 94% correlation between frequency of alcohol ta king and time to alcoholism. This indicates that the model/distribution fits the data quite well. Frequency of alcohol taking per week is highly correlated to in cubation period
3. Adjusted R^2 =0.875, this indicates that adding another variable to the frequenc y of alcohol taking per week will produce better prediction since adjusted R^2 is high 88%
4. The value of AIC is 2.966535. This points that the model fit is good.
5. BIC has a value of 2.981256 and HQIC giving 2.97247

The graphs and fit indices all gave good fit for the frequency of alcohol drinking per w eek in the theorised B-S distribution.

4.2.4 Hazard function of B-S distribution

The hazard function, Figure 4.15 confirms that cdf and pdf is actually B-S distribution.

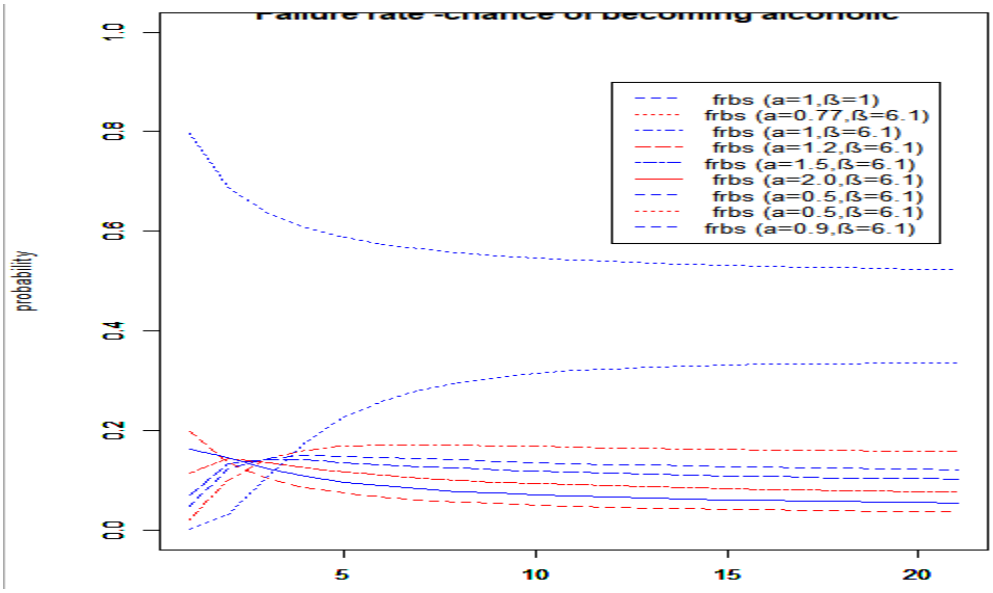


Figure 4.15 hazard curve of the BS distribution for the indicated values.

The hazard curve above confirms the characteristic of B-S distribution. B-S distribution is indeed the model suited to describe time from onset of alcohol taking to becoming alcoholic.

4.3. Hazard of Becoming Alcoholic

The two main options used in reporting the results of the hazards model are either numerical or graphical. Since the proportionality condition/assumption was violated then the results were mainly presented graphically.

Parametric hazard rate was first briefly presented in section 4.2.4 based on B-S distribution. Figure 4.16 shows the parametric hazard rate when alpha and beta values are the MLE values. The hazard curve is upside down (n-shaped).

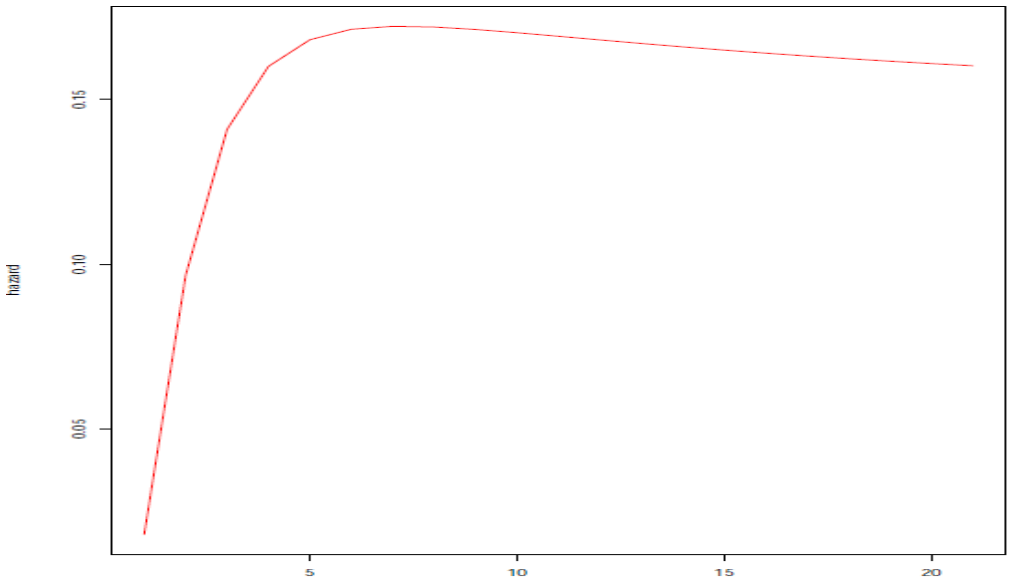


Figure 4.16. Hazard rate of getting AUD as frequency of drinking per week increases

The hazard rate is increasing as frequency of alcohol taking per week increases up to the median or a value near it. Thereafter it stabilises and then declines gradually. The associated failure rate is given in Figure 4.17. Graph indicates that as the frequency of alcohol taking per week increases so does the risk of becoming alcoholic. The failure rate is an increasing curve.

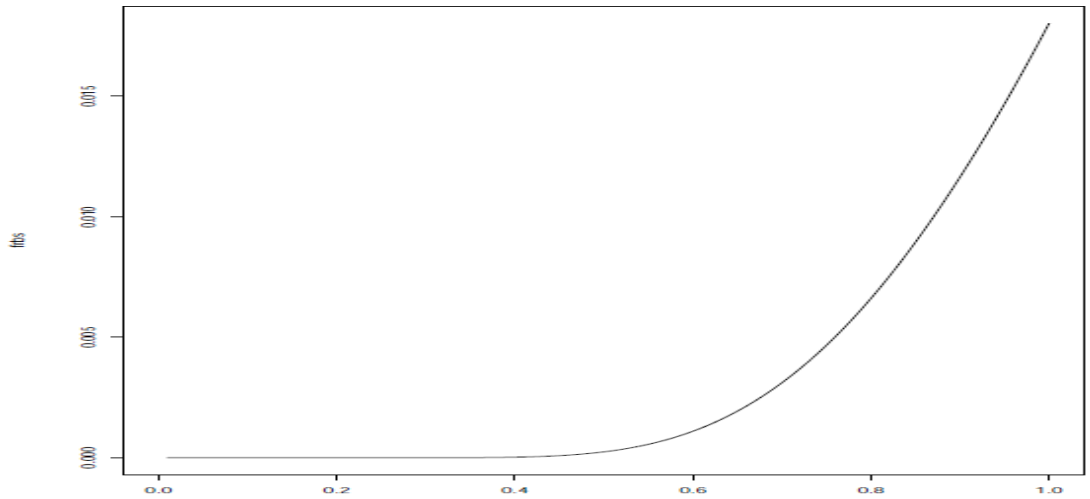


Figure 4.17 Failure rate

4.3.1 Probability of becoming alcoholic.

Probability of acquiring alcoholism was evaluated based on frequency of drinking per week and on the time-period over which one has been drinking. These are given in Table 4.4 and Table 4.5 respectively.

Table 4.5. Probability of acquiring AUD with frequency of drinking per week.

Frequency per week	1	2	3	4	5	7	10	12	15	20
Probability of getting AUD	0.003	0.06	0.167	0.285	0.393	0.569	0.742	0.816	0.889	0.951
Prob in %	0.31	5.96	16.7	28.5	39.3	56.9	74.2	81.6	88.9	95.1

The higher the frequency of drinking the higher the probability of getting AUD. An increase in frequency of drinking from once to twice a week raises the chance of becoming alcoholic by 20 times.

Table 4.6. Changes in probability of acquiring AUD with time.

Time since onset (years)	1	3	5	10	15	20	30	40	50	55
Probability of getting AUD	0.00006	0.025	0.096	0.288	0.437	0.548	0.699	0.793	0.855	0.878
Prob in %	0.0059	2.53	9.55	28.8	43.7	54.8	69.9	79.3	85.5	87.8

While the risk is increasing with time since onset of alcohol taking. While continuing to drink heavily from year one to year three increases the risk of becoming alcoholic by over 400 times.

4.3.2 Non-Parametric hazard rate based on measured variables

Estimated hazard function comparing hazard between males and females

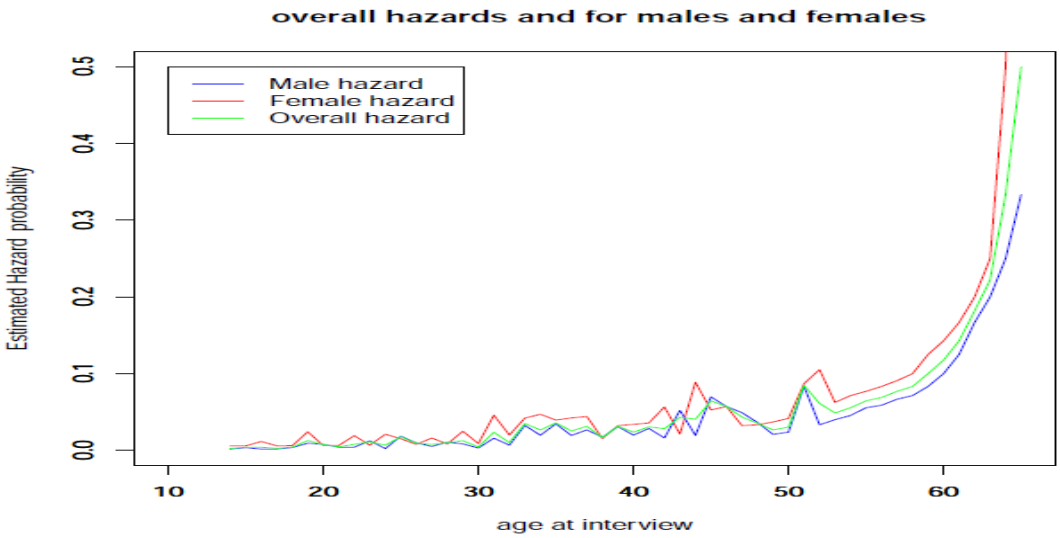


Figure 4.18: Gender based hazard of onset of alcoholism

From hazard plots based on life tables in Figure 4.18, it was noted that:

- The hazard for alcoholism was higher for females than for males.
- For both males and females the hazard of alcoholism had an increasing trend.
- The hazard obtained by combining males and females was sandwiched between that of males (highest overall) and that of females (lowest).

Survival curves were plotted to compare survival of moderate male and female alcohol takers before becoming alcoholic. Survival curves plotted using life table was given in Figure 4.19.

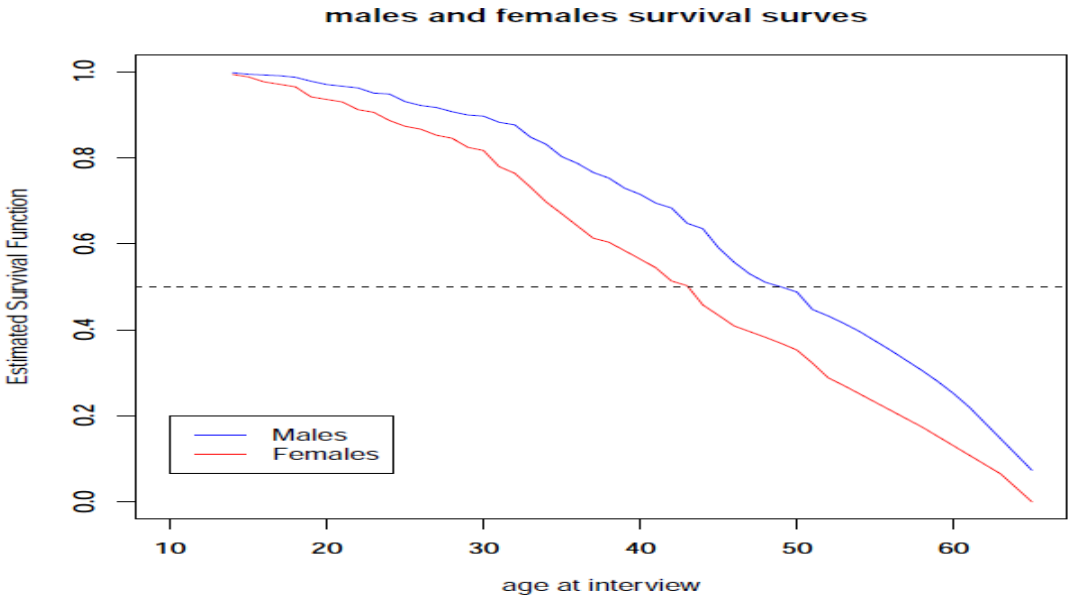


Figure 4.19: Survival plot for persons who are consuming alcoholic drinks.

It was noted that the survival probabilities for males is higher than for females, which means that at any given age the percentage of males who survived was higher than the percentage of females who survived. Thus, males who are taking alcohol stay for longer before becoming alcoholic compared to females who are drinking.

4.3.3 Hazard and survival probabilities using the life tables

Combined hazard and survival graph (Figure 4.20) show that hazard is generally increasing with age, survival declines with age.

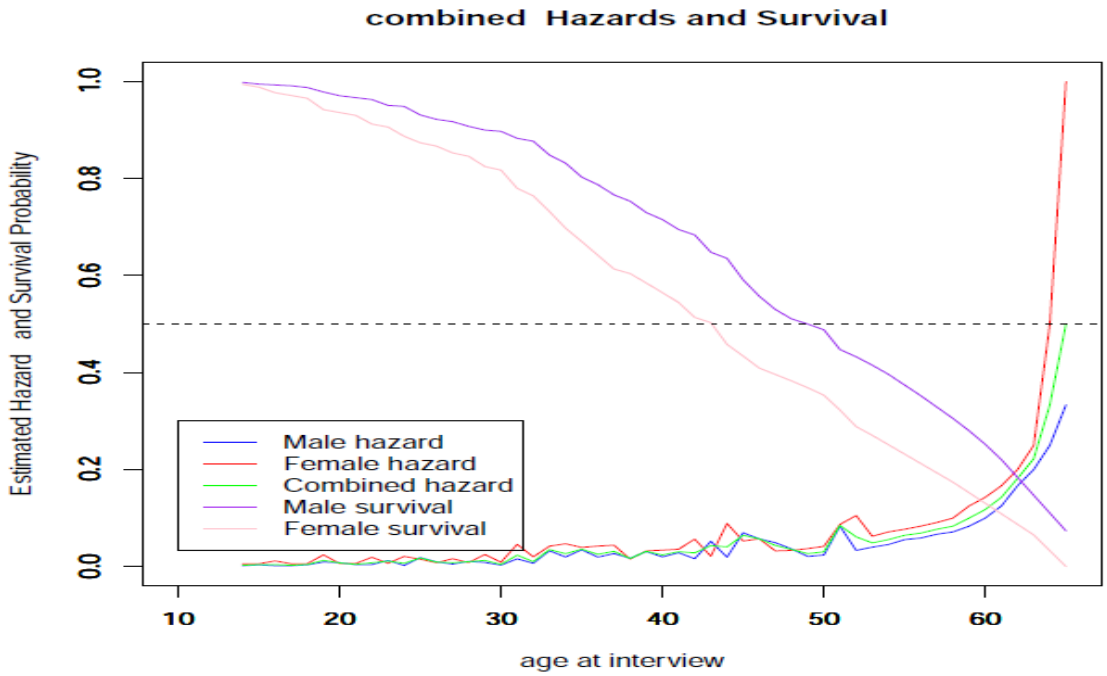


Figure 4.20: Hazard and survival curves

4.3.4 Logit-Based models for Hazard Function

By converting probabilities to odds, a scale which is bounded between 0 and 1 was transformed into a range from 0 to infinity. The resulting plot was shown in Figure 4.21. The curve for odds for females was generally higher than the odds for males but still cross at some ages.

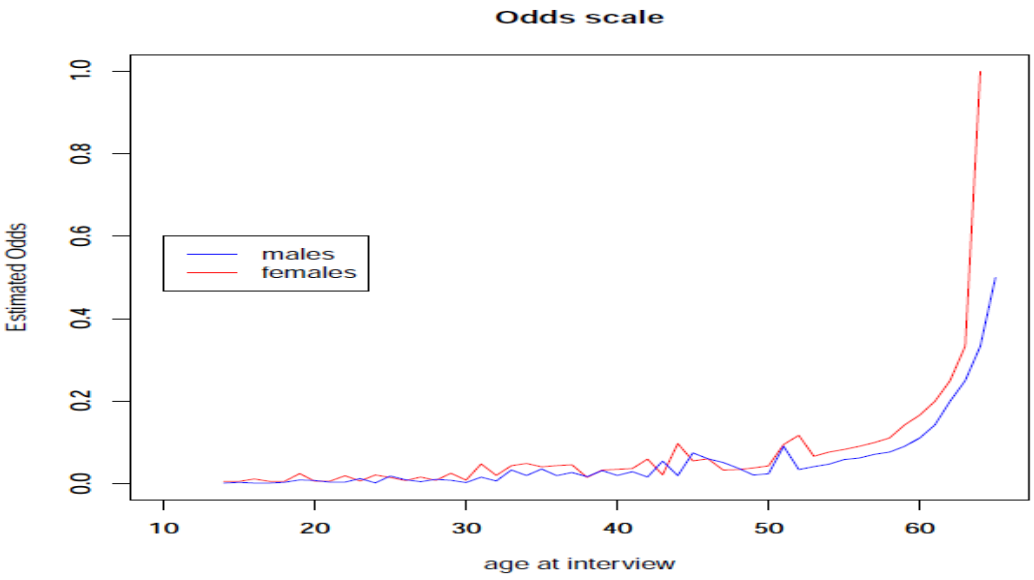


Figure 4.21: Graph of odds

Plotting using Logit scale increases the range from negative infinity to positive infinity. The resultant graph is in figure 4.22

$$Logit(p) = \log\left(\frac{p}{1-p}\right) \qquad 4.3$$

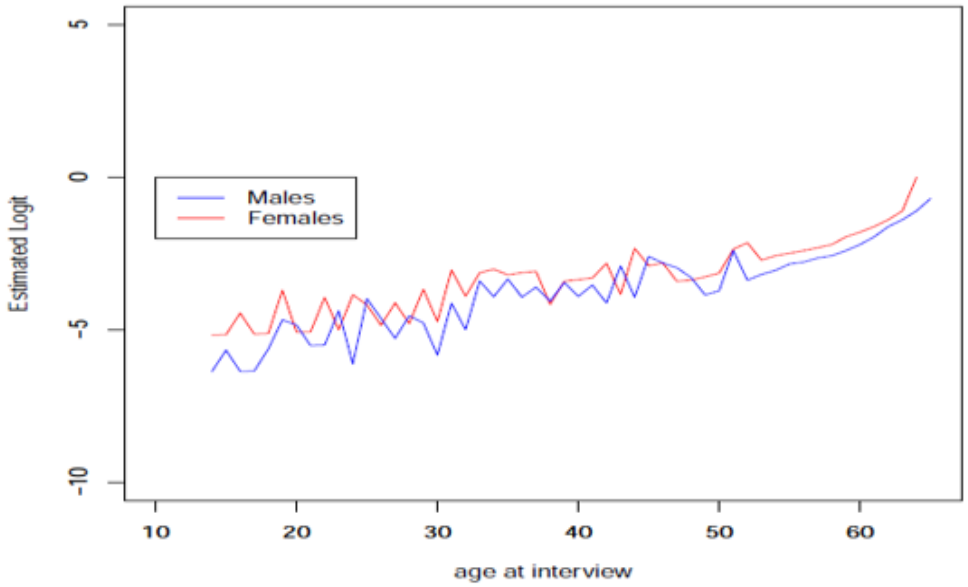


Figure 4.22: Logit curves for the hazard

The behaviour of the non-parametric hazard rate with respect to factors and changes in time was given in Figure 4.23. Even though it was not smooth it was comparable to the parametric hazard rate in Figure 4.16

The fact that logit curves crossed one another at some points was a clear indication that a given covariate did change over time. This was a violation of major assumption of the Cox proportional hazards model. Since this assumption was violated, the simple Cox model is invalid, and more sophisticated analyses are required.

When duration or time since onset of alcohol taking that is how long one has been drin king alcohol was considered it was found that alcoholism was significantly influenced

by the duration, p <.1%. However, drinking for less than one year was not significant a s illustrated in the Appendix H.

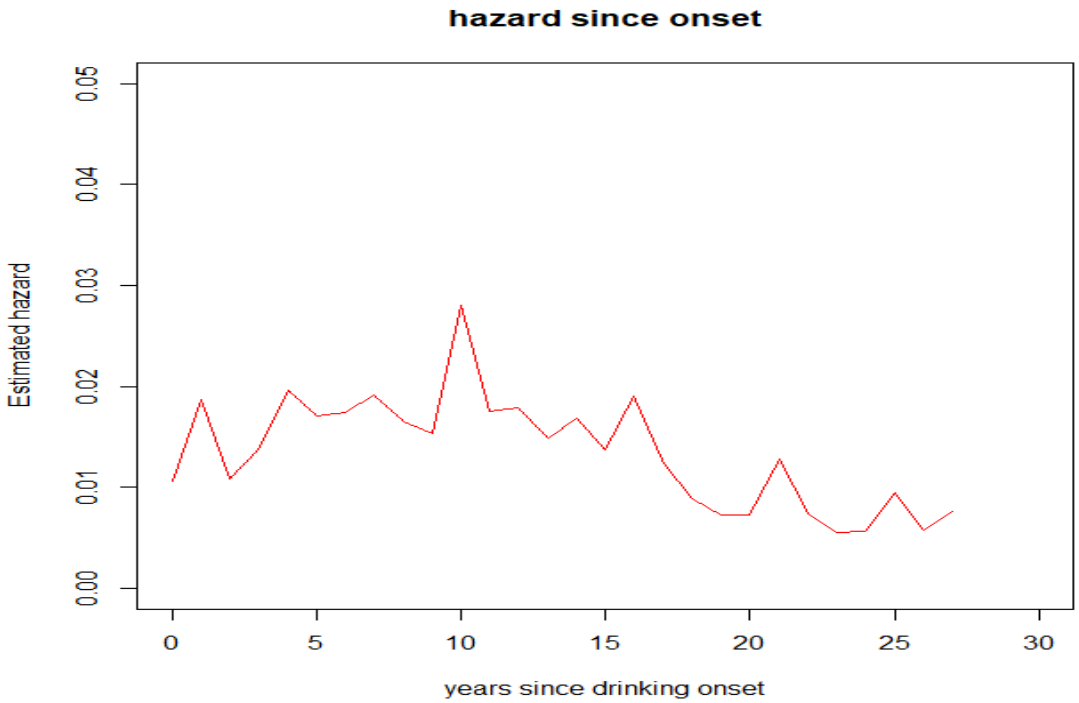


Figure 4.23. Graph of hazard of alcoholism since onset of alcohol taking

4.3.5 Effect of age at onset of alcohol consumption on hazard function

Age at onset is one of the key predictors of alcoholism mentioned in literature. Its effe cts on the hazard function were assessed and reported as follows: By using age at ons et as the sole predictor it was notable that starting to consume alcohol under age of 12 years, and ages between 13 to15 years and 16-19 years one was most likely to get into problems with alcohol use, p< .1% .This is illustrated in Table 4.7. Note that ageonset refers to age at onset of alcohol taking

Table 4.7: Parameter estimates for different ages for onset of alcohol taking.

Factor	Estimate	Std. Error	z value	Pr(>|z|)
(ageonset < 12)	-0.7538	0.1621	-4.651	0.00000329607 ***
(ageonset 13-15)	-0.9337	0.1563	-5.972	0.00000000234 ***
(ageonset 16-19)	-0.6931	0.1987	-3.489	0.000485 ***
(ageonset 20-24)	-0.4055	0.2440	-1.662	0.096530
(ageonset 25-30)	-0.3973	0.2019	-1.967	0.049145 *
(ageonset 31-35)	-0.6190	0.3315	-1.867	0.061845 .
(ageonset >35)	-0.8109	0.3005	-2.699	0.6956 **

.

4.3.6 Effect of Gender on hazard function

When gender was used as the baseline it was noticeable that being diagnosed as alcoh olic was significant for male p< 0.1% as shown in Table 4.8

Table 4.8. Parameter estimates for gender

Factor	Estimate	Std. Error	z value	Pr(>|z|)
Gender (Female)0	0.1245	0.1506	0.826	0.409
Gender (Male)1	-0.9951	0.0937	-10.620	<2e-16 ***

The results of comparing hazards showed that females had a higher baseline hazard than males as shown in Figure 4.24. When factors were added one after the other to the baseline model the lines, bumps appear on both lines, Figure 4.25. Figure 4.26 shows that distortions are largest when all factors are added into the baseline model.

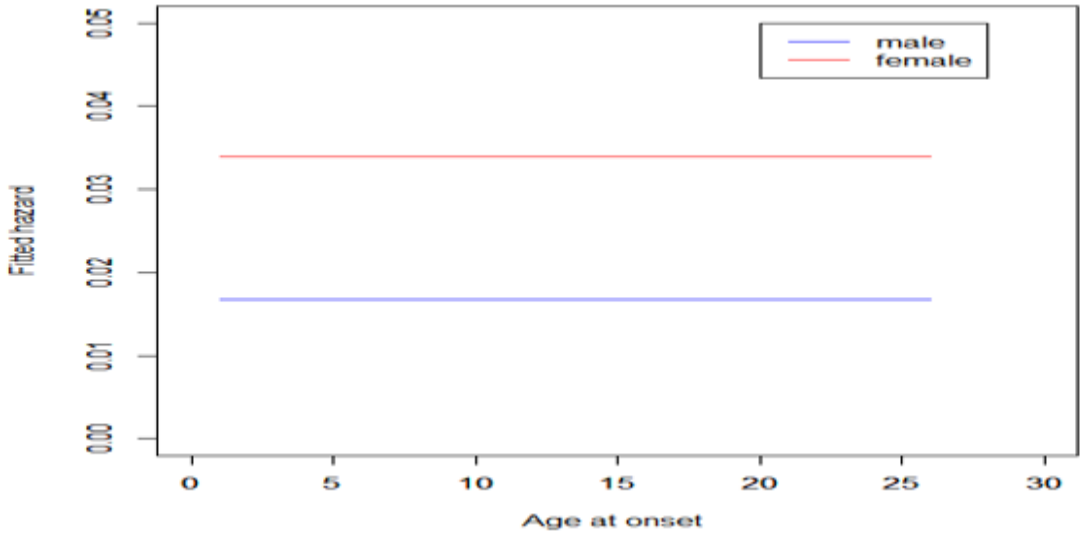


Figure 4.24. Comparing hazard between males and females

The bumps are more pronounced in females than in males. At a given age of onset some factors increased hazard while others decreased it. A factor had different effect on an individual depending on the age at onset.

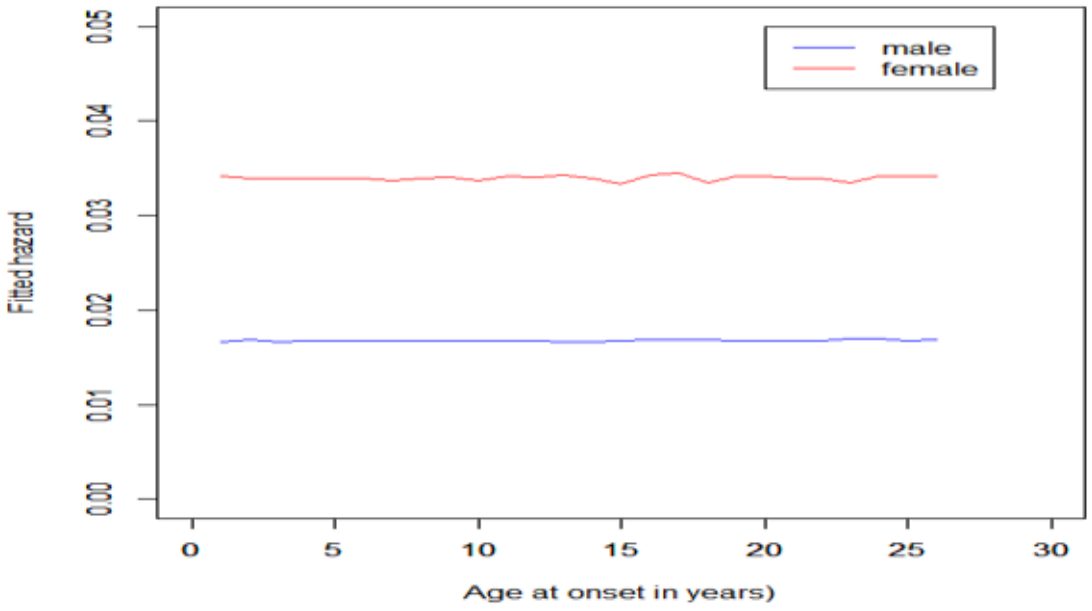


Figure 4.25 Effect of adding socio-cultural factor to the baseline model.

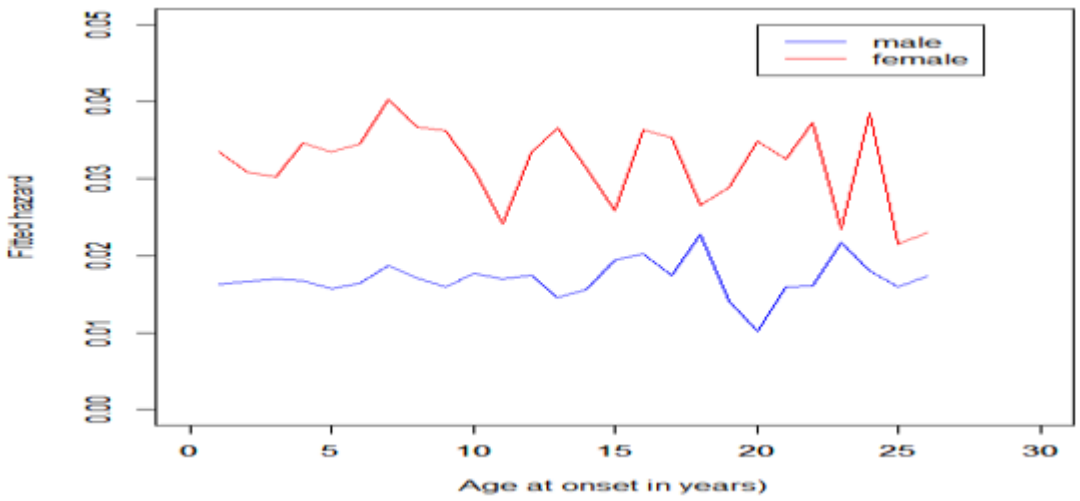


Figure 4.26 Effect on hazard after adding all the factors to age at onset (baseline)

When the baseline was changed to time since onset (the duration an individual has been drinking) the pattern changed as shown in figure 4.27. The dotted line is the baseline.

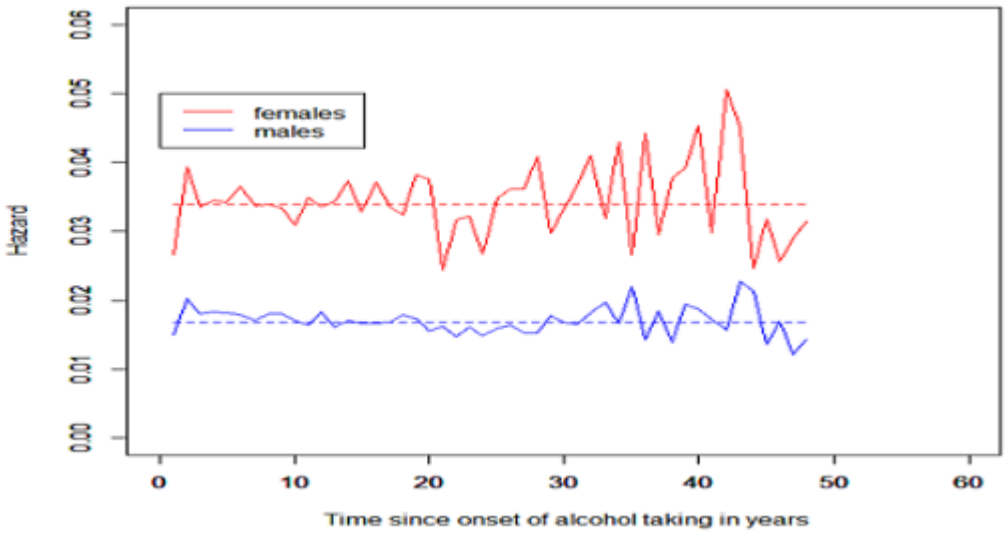


Figure 4.27. Effect on hazard after adding all the factors to baseline model

The effect of factors on hazard are more pronounced for those who have been drinking over long period of time as observed from Figure 4.27. In addition, females were more affected than males. Repeating the same process with one factor at a time the results produced patterns shown in figures 1 to 8 the appendix C.

Economic status, personality, drinking habits and patterns and peer influence had the greatest impact on hazard in the long run. While environment, socio-cultural, family and family attention and age at onset had low impact on hazard over long term.

4.3.7 Models based on measureable variables.

In this section results of model fitting are presented. Three baseline models characteristics were discussed. These are based on (a) the period since onset of alcohol taking (duration of drinking), (b) age in the year of diagnosis and (c) age of onset of alcohol consumption.

(a) In the model where the period since onset of alcohol taking (duration of drinking) was taken as the baseline (Model A) it was notable that: There was no detectable statistical significance in the first year of alcohol taking with respect to one becoming alcoholic. The effect of duration on alcoholism was significant from two years since onset of alcohol taking all through to the 27^{th} year ($p<$ .1%). (27 years was the longest duration before onset of alcoholism). This statistical significance rose higher five years after onset.

(b) When age in the year of diagnosis was used as the baseline model (Model B) it was noted that; Ages 14years through to 37years were highly significant ($p<$ .1%). Thus being diagnosed as alcoholic was highly likely for persons of the ages 14-37. Those aged below 14 years and those over 40 years old had little chance of being detected as alcoholic at, p=.1% significance level. See Appendix G2

(c) When age of onset of alcohol consumption was used in the baseline model (Model C); high statistical significance ($p<$ .1%) started right from age of 10 years all the way to the age of 31 years was noted.

(i) Effect of adding all measurable variables to the baseline model (Model A).

When all variables were added to the baseline model (Model A), statistical significance changed. Variables such as happiness in the house, type of house where one lives, number of alcohol selling points in the vicinity were fairly significant (p<5%). Gender was extremely significant (p< .1%). Other variables such as knowledge about dangers of excessive alcohol consumption and ignoring them, frequency of drinking, ones' religion and perception of alcohol raising ones' confidence were less significant (p> .1%). AIC in model A^* declined compared to that of model A as shown in Table 4.8.

Table 4.9. Comparing fit indices between baseline model and model A^*

Measures of fit	Base- line model (Model A)	Model A^*
Residual deviance	2000.3	1912.2
AIC	2096.3	2086.2.
Degrees of freedom	11386	11347

The R program version of Model A^* is represented below. The acronyms used in the model are explained in Appendix E2

Model A^* glm(formula = event ~ factor(PERIOD) - 1 + act_rel + adalc + a lc_hme + ageonset + alc_law + am_hse + app_frd + bsf_drn + cel_soc + com_ cer + dang_ex + emp_bis + fam_gath + fe_risk + fred + freq + Gender + gene + hap_hse + hero_ex + high_edc + hse_type + job_ex + joints + las_drnk + lov _feel + mar_sta + neighb + p_educ + pa_liv + peer_dr + quanty + race + relig n + self_con + sib_drn + sport + stre_wrk + typealc, family = binomial(logit), data = fox)

(ii) Effect of age at onset of alcohol taking (Model C)

When all variables were added to the baseline model (Model C), only ages 10, 11 and 16 years were statistically significant (p< .5%). All the other ages gave no statistical significance with respect to experiencing alcoholism in the presence of other variables. Variables such as being active in religious matters, ones' religion and gender are significant in predicting alcoholism, (p<.1%). Gender and being active in religious activity were highly significant (p< .1%). AIC in model C* declined compared to that of model C as shown in Table 4.9.

Table 4.10. Comparing fit indices between baseline model and model C*

Measures of fit	Base- line model (Model C)	Model C*
Residual deviance	2260.5	2166
AIC	2314.5	2296
Degrees of freedom	11407	11347

The R program version of Model C* is represented below. The acronyms used in the model are explained in Appendix H

Model C* glm(formula = event ~ factor(years.onset) - 1 + act_rel + adalc + alc_h me + ageonset + alc_law + am_hse + app_frd + bsf_drn + cel_soc + com_cer + da ng_ex + emp_bis + fam_gath + fe_risk + fred + freq + Gender + gene + high_edc + hero_ex + hap_hse + hse_type + job_ex + joints + las_drnk + lov_feel + mar_sta + neighb + p_educ + peer_dr + pa_liv + quanty + race + relign + self_con + sib_d rn + sport + stre_wrk + typealc, family = binomial(logit), data = fox)

(iii) Most parsimonious model (Model A**).

When model A* was subjected to stepwise with backward shift model selection method using the Rcmdr package in RStudio applying AIC criterion, the final model selected contained the following measureable variables: Gender, ignoring dangers of excessive drinking, excessive alcohol taking running in the family, education level of the alcohol taker, religion, stress at work place, what parents do for a living, amount of alcohol ordinarily drunk per session and love for feeling of being drunk. The coefficients in model A** are listed in Table 4.12. The AIC reduced to 890.7 as shown in Table 4.11.

Table 4.11. Comparing the model fit indices for the three models

Measures of fit	Model A	Model A*	Model A**
Residual deviance	2000.3	1912.2	870.7
AIC	2096.3	2086.2.	890.7
Degrees of freedom	11386	11347	754

Table 4.12 Coefficients for variables in model A**

Variable	(Intercept)	dang_ex	Gender	gene	high_edc	lov_feel	pa_liv	quanty	relign	stre_wrk
coefficient	1.295	-0.418	-1.248	-0.395	-0.224	-0.367	-0.407	0.417	0.166	0.406

The acronyms are as explained in Appendix E2

4.3.8 Models based on Factors

These are models based on one or combination of nine factors that were generated from the measurable variables discussed in section 4.1, namely: gender, peer pressure, personality, family and family attention, economic status, socio-cultural, drinking habits and patterns, age at onset of alcohol taking and environment/structural setting.

i. Baseline model (Model D)

The baseline model consisted of period (duration from onset of alcohol taking to diagnosis) in years as the only factor. All the years, starting from one year after onset of alcohol taking were significant, ($p < .1\%$).

ii. Model D*. Adding gender to baseline model.

Gender was highly significant ($p< .1\%$), period 0 is still nonsignificant, while period 1 now becomes less significant ($p< 5\%$).

iii. Model D**. Full model: All factors fitted in the model.

Gender and personality were significant as were periods except for period 0 and 1 ($p< .05$)

AIC declined as factors were added to the baseline model. It was lowest in full model. Change in the AIC for models Model D, Model D* and Model D** was given in Table 4.13.

Table 4.13. Comparing the model fit indices for the three models

Measures of fit	Model D	Model D*	Model D**
Residual deviance	2254.6	2226.5	1360.6
AIC	2354.6	2328.5	1588.6
Degrees of freedom	11705	11704	11641

The coefficients of the risk factors were presented in Table 4.13.

Table 4.14: Coefficients of risk factors of alcoholism

Factor	coefficients	Exponentiated coefficient	significance
Gender	-0.90547	4.043528e-01	0.001 ***
Age at onset	-0.00036	9.996400e-01	0.45
Peer influence	0.03151	1.032008e+00	0.23
personality	-0.17811	8.368504e-01	0.05 *
Socio-cultural	0.10544	1.111194e+00	0.14
Economic-status	-0.11733	8.892891e-01	0.07
Environ/Structural	0.10581	1.111609e+00	0.33
Family attention	-0.08603	9.175645e-01	0.17
Drinking habit and pattern	-0.02543	9.748896e-01	0.92

4.4 Projections of alcoholism incidences in Kenya

Secondary data on cases of morbidity from alcoholism was cleaned from errors of commission and omission were corrected after consultation and then aggregated, see Table A.1 and Table A.2 in Appendix A. Morbidity data contained cases of the disease conditions which are by definition alcohol-attributable (AAF = 100%) as listed in Appendix A.

4.4.1 Alcoholism incidences reconstructed using logistic model

The glm model that was applied is as in equation 3.18. The model with the lowest AIC (45.2) and residual deviance (124.1) was,

$$\hat{\mu} = 13.362 - 1.382 * t + 0.00196 * exp(3.5 * t - d) - 0.00033 * t * exp(3.5 * t - d) \quad 4.4$$

*w*here $d = t\text{^}.1$

The predicted values of incidences of alcoholism are given in Table 4.15.

Table 4.15: Predicted and observed incidences of alcoholism from 2014 to September 2019.

Year	Observed (y)	Projected ($\hat{\mu}$)	$y - \hat{\mu}$	$(y - \hat{\mu})^2$	$\frac{(y - \hat{\mu})^2}{\hat{\mu}}$
2014	9	11.99996	-2.99996	8.99976	0.749983
2015	19	11.09008	7.90992	62.56683	5.641694
2016	14	20.9297	-6.9297	48.02074	2.294383
2017	261	258.871	2.12896	4.532471	0.017509
2018	4091	4091.159	-0.15869	0.025183	6.16E-06
2019	2237[a]	2237.898	-0.898	0.806404	0.00036
Total	**6631**	**6632**			**8.7039**

[a] This value relates to the period ending September 2019. No adjustments were made to account for the remaining quarter of the year.

Comparing prediction incidences to the reported alcoholic cases gave a deviation of 8.70. χ^2(6, N=6631)=8.70, p <.05, thus the predicted of the number of alcoholics (6632) did not differ significantly from the reported cases of diagnosed alcoholics (6631). The curves fitted using observed and projected values is given in Figure 4.28. The fit was quite good.

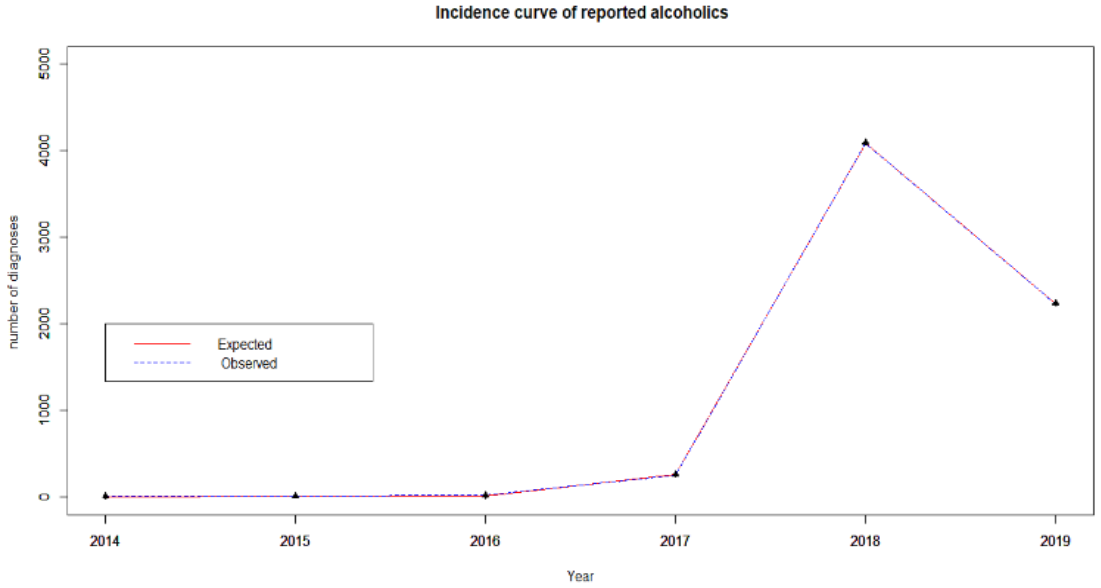


Figure 4.28: Observed and predicted curve of incidences of alcoholism

Short term projections (based on annual cases) of incidences expected to be diagnosed and admitted in various health centres locally gave negative value. This was unexpected. The remedy was to base projections on quarterly incidences (discussed in section 4.4.3). Back-projection was done using discrete approximation on the values of diagnosis to get the corresponding cases of alcohol user initiates.

4.4.2 Back-projected alcohol using initiates.

This was based on the operationalisation of equations 3.22, 3.23 and 3.24. First, probability matrix $[(f(j,i))]$ was determined using equation 3.24. The values of $\alpha \; and \; \beta$ used were alpha = 0.93617 and beta = 10.8877. The resultant matrix, $[(f(j,i))]$ was an upper triangular matrix with elements zero in the lower triangle (below the leading diagonal). It was inverted and multiplied with the column matrix of the number of alcoholics reported between 2014 to September 2019. These risky alcohol users were expected to give rise to the alcoholics projected. Two predictions give negative values. This is one drawback of back-calculation method.

4.4.3 Incidences and Projections of Alcoholics for each quarter year

The logistic model with the least AIC and lowest Residual deviance was

$$\hat{\mu} \sim t * (1 + \ + exp(-(7/3) * t + \ (t\text{^}0.25)))$$

AIC = 72.98 and Residual Deviance =3313, Null deviance =166100

The data on morbidity (Table.A2 in the appendix) was fitted as a glm model using the formula described in equation 3.18. The resultant equation is shown below.

$$\hat{\mu} = 3457 - 123.2t + 7.101 * exp\left(-\left(\frac{7}{3}\right) * \ \mathrm{t} + (\mathrm{t}^{0.25})\right) - 4.224\mathrm{e} + 19 * \mathrm{t}$$

$$* \ \exp(-(7/3) \ * \ \mathrm{t} + (\mathrm{t}\text{^}0.25))$$

On fitting the data in the model the predictions were compared to the reported incidences of alcoholics. The fitted values were compared to the reported ones in Table 4.16.

Table 4.16 Predicted and observed annual incidences of alcoholism

Year	2018 :Q1	2018 :Q2	2018 :Q3	2018 :Q4	2019 :Q1	2019 :Q2	2019 :Q3	**Total**
Observed (y)	1005	1014	1070	1002	835	787	607	**6320**
Projected ($\hat{\mu}$)	1002	1013	1075	987	869	747	623	**6315**
$y - \hat{\mu}$	3	1	-5	15	-34	40	-16	
$\frac{y - \hat{\mu}}{\sqrt{\hat{\mu}}}$	0.095	0.031	-0.153	0.4775	-1.153	1.4635	-0.641	**0.12**
$\frac{(y - \hat{\mu})^2}{\hat{\mu}}$	0.009	0.001	0.0233	0.228	1.33	2.142	0.4109	**4.1442**
$Time$	17	18	19	20	21	22	23	

The Pearson statistic is 4.1442. Now $\chi^2(6, N=6320) = 4.14$, $p < .05$, indicating that the predicted number of alcoholics is not significantly differently from the number of observed or alcoholics diagnosed.

The projected and observed alcoholism incidences were plotted on the same axes as given in Figure 4.29.

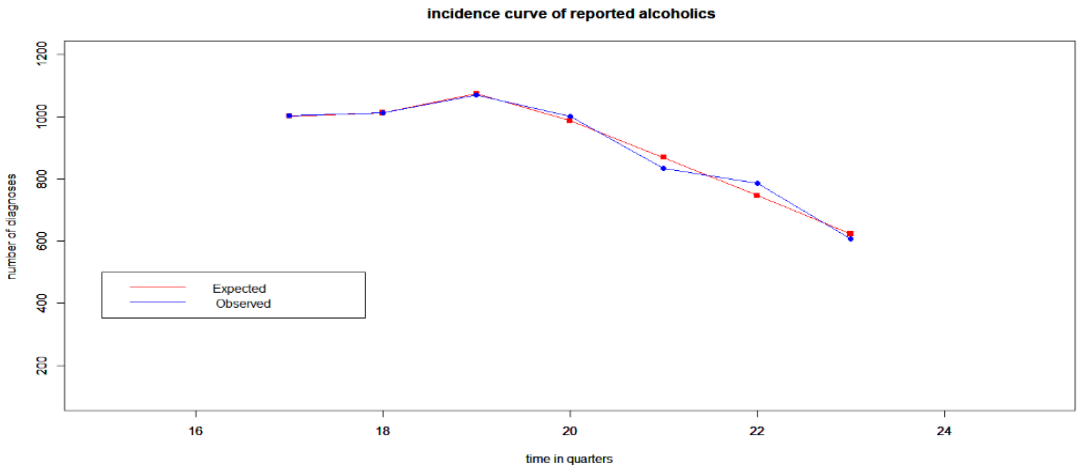


Figure 4.29: Quarterly observed and predicted curve of incidences of alcoholism

Using the results of curve fitting short term projections were made up to June 2020.

Table 4.17: Projections of alcoholics between Jan 2018 and June 2020.

Year	2018 Q1	2018 Q2	2018 Q3	2018 Q4	2019 Q1	2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	**Total**
Back-projection	1002	1013	1075	987	869	747	623	500	377	254	**7447**

Between 2018 and mid 2020 more than 7000 alcoholics are expected to have been reported diagnosed and admitted in various health centres locally.

Back-projection was done using discrete approximation on the values of diagnosis to get the corresponding cases of alcohol user initiates. First, probability matrix $[(f(j,i))]$ was determined using equation 3.24. The values of α *and* β were determined modified

moment method. Alpha = 0.20433 and beta = 15.34. The resultant matrix was an upper triangular matrix with elements zero in the lower triangle below the leading diagonal. The number of alcoholics predicted from 2018:Q1 to 2019:Q3 is given in Table 4.18.

Table 4.18: Back-projection of new risky alcohol users

Year	2018 Q1	2018 Q2	2018 Q3	2018 Q4	2019 Q1	2019 Q2	2019 Q3	**Total**
Back-projection	-16	3575	4259	3710	3150	2654	2192	**19525**

A total of approximately 19,500 persons had been inducted into risky alcohol taking between January 2018 and September 2019. Projection for first quarter 2018 is negative which, is one of the shortcomings of back-projection. However the cumulative total give a fair view of the evolution of new alcohol users. There were over twenty two thousands new risky alcohol users are expected by end of June 2020 as given in Table 4.19.

Table 4.19: Back-projection of new risky alcohol users up to June 2020.

Year	2018 Q1	2018 Q2	2018 Q3	2018 Q4	2019 Q1	2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	**Total**
Back-projection	-16	3485	4169	3620	3060	2564	2102	1670	1251	840	**22745**

4.5 Discussions

Path analysis showed that nine risk factors (drinking habits and patterns, age at onset, gender, personality, peer influence, family and familiar attention, economic, socio-cultural and environment) caused alcoholism as illustrated by favourable fit indices. Some factors for example, gender acted directly and also through drinking habit and pattern. This is in agreement with work by several researchers. However, there was no

agreement on the combination of risk factors that cause alcoholism and the use SEM in their analysis was also limited. For example, without use of SEM ICAP (2009) described how social-economic, environment determinants, individual characteristics and genetic disposition interacted to influence drinking choices which in turn formed drinking patterns which gave produced drinking outcome. Drinking outcome again informed patterns and choices. Tripathy et al. (2018) used SEM to identify the risk factors though in diabetes. They identified five variables including alcohol use as the key drivers of diabetes.

Time from onset of alcohol taking to diagnosis with alcoholism was modelled using B-S distribution (α =0.77 [CI: 0.68, 0.85] and β = 6.13 [CI: 5.44, 6.83]. The use of B-S distribution was justified based mechanism by which data on alcoholism is generated, Becker (1990). There was evidence of damage to cells and body tissues due to prolonged excessive alcohol use. The use of B-S distribution in other failure time cases have been illustrated in the literature review. There was no evidence of study of time to alcoholism using any failure time model by the time of this write up. Time to alcoholism has mainly been survey based.

Probability of becoming alcoholic was evaluated as a function of frequency of drinking per week and later as a function of time since onset of risky alcohol taking. In both cases probability increased both with frequency and with time since onset.

The parametric hazard function of alcoholism obtained was n-shaped. The non-parametric hazards model based on measured variables violated proportional condition in Cox proportional hazards model. Indicating some variables changed with time. More complex techniques were needed to do analysis. Factor resultant based model described

by equation 3.5 showed that different factors affected the hazard of becoming alcoholic differently and females had higher hazard of becoming alcoholic than males.

Gender and personality were highly significant in influencing alcoholism. Hazards rate of females was higher than that of males. Findings on gender difference in alcoholism is well documented with biological and psychological reasons being used to explain this difference. For example Horihan (2014), Walter et al. (2005) among others found gender difference on effects of alcohol.

The glm predicted 6632 incidences which did not differ significantly from 6631 reported alcoholic cases χ^2(6, N=6631) =8.70, p < .05. Back-calculation method was applied in making projections, estimating the trend patterns of incidences of both alcoholism cases and onset of risky alcohol taking using the morbidity data. Locally, these estimates have been survey based. For example, NACADA in undated survey article on alcohol and drug use in students found a 23.4% prevalence rate, NACADA (n.d.). More recently NACADA estimated that there were 2.65 million drug users in Kenya, (Wambui, 2015).

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.0 Summary

The main purpose of this study was to analyse the structural relationship between alcoholism and its risk factors and model alcoholism as a non-communicable disease. The specific objectives were to: Analyse the structural relationship between alcoholism and its risk factors, model incubation period of alcoholism, evaluate hazard rate of alcoholism and make projections for incidences of alcoholism in Kenya. A cross-sectional study which applied multistage stratified sampling method was done in seven counties and the results were discussed.

5.1 Conclusions

The nine risk factors (gender, personality, age at onset, peer influence, family and family attention, economic status, social-cultural, environment, and drinking habits and patterns) fitted fairly well in the hierarchical SEM alcoholism model as indicated by the indices that were within their respective favourable ranges. RMSEA and SRMR were within acceptable limits; while TLI, and CFI were slightly below their cut-offs.

Factors describing personality included love of feeling drunk, drinking promoted self-confidence, drinking excess alcohol showing one is a hero, and ignoring warning about dangers of excessive drinking. Family and family attention was linked to parents/guardian/siblings drinking alcohol, marital status and happiness, peace, fun or joy at home. Socio-cultural was linked to involvement in sports, race, involvement in religious activities and community having occasions/ceremonies where alcohol taking by all was permitted.

Based on biophysical foundation (where the process of alcohol taking to alcoholism mirrors the way B-S distribution is derived) and the goodness of fit indices, the B-S

model fitted the data well with adjusted R^2 =0.88. In addition parametric hazard function was an upside-down curve confirming that incubation period distribution was best modelled using B-S distribution.

Probability of becoming alcoholic increased with increase in frequency of drinking per week and with time the duration one had been taking alcohol. While gender and personality had the largest influence on hazard rate (p< .1%), hazard rate for males was lower than that of females. Hazard for males and females becoming alcoholic varied with age and each was affected by factors differently.

Logistic model based projections were not significantly different from observed cases χ^2(6, N=6631) =8.70, p < .05. It was Back projected that over 19,000 persons were expected to have initiated alcohol use between January 2018 and September 2019, rising to 22000 by mid- 2020. Availability of more accurate data especially on morbidity, knowledge of reporting delays would help to generate more accurate predictions.

5.2 Recommendations:

Further research

1. Interaction, mediation and confounding effects among the nine risk factors discussed should be investigated.
2. Research involving spatial regression analysis be done to get a picture of how alcoholism morbidity and mortality vary in different counties/geographical regions. Spatial modelling could be done to show the regional variation and trends.

Policy

1. Focus on sensitizing females to cut the volumes of alcohol taken and also initiate programs targeting to reduce the number of young men starting to take alcohol.
2. Use the information on risk factors to design factor based prevention programs and apply projections to make provisions for the future health needs.

REFERENCES.

Ahrens, W. & Pigeot, I. (Eds.) (2006). *Handbook of Epidemiology*. Berlin: Springer.

Aksu, G., Eser, T., & Guzeller, C.O. (2017). Analysis of Maths Literacy Performances of Students with Hierarchical Linear Modeling (HLM): The Case of PISA 2012 Turkey. *Education and Science 2017, Early Release, 1-20*. DOI: 10.15390/EB.2017.6956.

Alac (2012) Alcohol Advisory Council of New Zealand /Health Promotion agency. *Alcohol and your health: Is your drinking okay?* Wellington. ALAC. Retrieved from https://www.hpa.org.nz

Allison, P. D. (1982). "*Discrete–time Methods for the Analysis of Event–Histories." In S. Leinhardt, Sociological Methodology 1982 ed*. (pp. 61-98) Beverly Hills: Sage. Accessed from https://statisticalhorizons.com

Allison, P.D. (1984). *Event History Analysis: Regression for Longitudinal Data.* Sage University Paper Series on Quantitative Application in the Social Sciences, series no. 07–041. Newbury Park: Sage.

Almasy, L (2003). Quantitative risk factors as indices of alcoholism susceptibility. *Annals of medicine, 35,* 337-343. DOI 10.1080/07853890310004903.

Anderson P, Baumberg B (2006). *Alcohol in Europe. A public health perspective*. London, Institute of Alcohol Studies. Retrieved from http://ec.europa.eu/health-eu/doc/alcoholineu_content_en.pdf

Artzrouni, M (2004) Back-calculation and projection of the HIV/AIDS epidemic among homosexual/bisexual men in three European countries: evaluation of past projections and updates allowing for treatment effects. *European Journal of Epidemiology; 19* (2): 171-179.

Ashley, M.J., Rehm, J., Bondy, S., Single, E., & Rankin, J. (2000). Beyond ischemic heart disease: are there other health benefits from drinking alcohol? *Journal of Contemporary Drug Problems 27*, 735–777.

ASTM D7774-12. (2013). *Standard Test Methods for Flexural Fatigue Properties of Plastics, D20.10.24, Ed*. West Conshohocken, PA: ASTM International, [Online]. Accessed from www.astm.org

Baer, D (2005). *Event History Models: A Brief Introduction.* , Department of Sociology, University of Victoria baer@uvic.ca

Baer, J.S., Kivlahan, D.R., Blume, A.W., McKnight, P., & Marlatt, G.A. (2001). Brief intervention for heavy-drinking college students: four year follow-up and natural history. *American Journal of Public Health, 91*(8), 1310-1316. doi: 10.2105/ajph.91.8.1310

Bagozzi, R.P. (1980). *Causal models in Marketing*. New York: Wiley.

Balakrishnan, N., Vilca, F., Santana, L., & Leiva, V. (2011). Estimation of Extreme Percentiles in Birnbaum-Saunders Distributions. *Computational Statistics and Data Analysis, 55*, 1665-1678.

Baliunas, D.O., Taylor, B.J., Irving, H., Roerecke, M. Patra, J., Mohapatra, S., and Rhem, J (2009). Alcohol as a risk factor for type 2 diabetes: A systematic review and meta-analysis. *Diabetes care. 32*, 2123-2132

Baridi, L.M.P.O. (2014). *Strategies of Regulating Alcoholism and Drug Abuse in Kenya's Line Ministries: A Case Study of The Ministry of Devolution And Planning* (Master's thesis University of Nairobi, Nairobi Kenya). Retrieved from www.nacada.go.ke

Becker, N G, Watson, L W and Carlin, J B (1991). A Method of Nonparametric Back-Projection and its Application to AIDS Data. *Statistics in Medicine 10*(10):1527–42 https://doi.org/10.1002/sim.4780101005

Becker, N.G & Marschner, I.C (2001). Advances in medical statistics arising from the AIDS epidemic. *Statistical Methods in Medical Research, 10*, 117–140

Becker, N.G (1990). *Statistics of Infectious Diseases*. Downloaded from https://iase-web.org/documents/papers/icots3/BOOK2/Plenary2.pdf?1402524955

Bellis, M.A. (2016). *The harm paradox, Harm caused to others and life course effects.* Downloaded from https://www.m-c-a.org.uk › documents ›2016_m_bellis

Benedict, B. (2007). Modelling Alcoholism as a Contagious Disease: How "Infected" Drinking Buddies Spread Problem Drinking. *Journal of SIAM News*, 40(3). Downloaded from https://scholar.google

Bentler, P.M. & Chou, C.P. (1987). Practical Issues in Structural Modelling. *Sociological methods and research, 16*(1), 78-117. Downloaded from www.googlescholar.com.

Berman, S.M. (1970). Occupation times of Stationary Gaussian processes. *Journal of Applied Probability, 7,* 721-733

Beseler, C.L., Aharonovich, E., Keyes, M.K., & Hasin, D.S.(2008). Adult Transition From At-Risk Drinking to Alcohol Dependence: The Relationship of Family History and Drinking Motives. *Journal of Alcohol Clinical and Experimental Research, 32*(4), 607–616. doi:10.1111/j.1530-0277.2008.00619.x

Bhunu, C.P. (2012). A mathematical analysis of alcoholism. *World Journal of Modelling and Simulation, 8*(2), 124-134.

Birnbaum, Z. W. and Saunders, S. C. (1969). Estimation for a family of life distributions with applications to fatigue, Journal of Applied Probability, 6, 328-347.

Birnbaum, Z. W. and Saunders, S. C. (1969a). A new family of life distribution. *Journal of Applied Probability*, *6*, 319-327.

Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., ... Weinstein, C. (2011). *The Global Economic Burden of Non-communicable Diseases.* Geneva: World Economic Forum. Retrieved from www.world-heart-federation.org

Borges, G., Bagge, C.L., Cherpitel, C.J., & Conner, K.R. (2016). A meta-analysis of acute use of alcohol and risk of suicide attempt. *Psychological Medicine, 47*(5), 949-957. DOI: https://doi.org/10.1017/S0033291716002841

Borges, G; Breslau. J., Orozco, R., Tancredi, D.T., Anderson, H., Aguilar-Gaxiola, S., Mora, M.M (2011). A cross-national study on Mexico-US migration, substance use and substance use disorders. *Drug and Alcohol Dependence*, 117(1), 16-23. https://doi.org/10.1016/j.drugalcdep.2010.12.022

Bourguignona, M., Silva, R. B., & Cordeiro, G.M. (2013). A new class of fatigue life distributions. *Journal of Statistical Computation and Simulation, 84*(12), 2619-2635. https://doi.org/10.1080/00949655.2013.799164 . Downloaded from www.tandfonline.com

Box-Steffensmeier, J.N., & Jones, B.S. (2004). *Event History Modelling: A Guide for Social Scientists.* New York: Cambridge University Press.

Bradshaw, D., Steyn, K., Levitt, N., & Nojilana, B. (2011). *Non-Communicable Diseases – A race against time*. Cape Town. Retrieved from https://books.google.com/books

Brenan, A., Purshouse, R., Rafia, R., Meier, P., Booth, A., O'Reilly, D., … Wong, R. (2008). *Independent review of the effects of alcohol pricing and promotion part B. results from Sheffield Alcohol Policy Model.* University of Sheffield. Sheffield. Downloaded from www.sheffield.ac.uk

Brinks, R. & Landwehr, S. (2014). Age- and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. *Journal of Theoretical Population Biology, 92,* 62–68. http://dx.doi.org/10.1016/j.tpb.2013.11.006.

Brookmeyer R. (1996). AIDS, epidemics, and statistics. *Biometrics* 52, 781–796. doi:10.2307/2533042

Brookmeyer, R & Gail, M.H. (1986). Minimum Size of the acquired Immuno-deficiency Syndrome (AIDS) epidemic in the United States. *The Lancet, ii:* 1320-1322

Brookmeyer, R & Gail, M.H. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *Journal of American Statistics Association. 83,* 301–308. doi:10.1080/01621459.1988. 10478599

Brookmeyer, R. & Damiano, A. (1989). Statistical methods for short-term projections of AIDS incidence. *Statistics in Medicine, 8*: 23–34.

Broström, G (2012). *Event History Analysis with R*. The R Series, Boca Raton, FL: Chapman & Hall/ CRC Press.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford Press.

Canchola, A.J., Stewart, S.L., Bernstein, L., West, D.W., Ross, R.K., Deapen, D., … Horn-Ross P.L.(n.d). *Cox Regression Using Different Time-Scales.* Northern California Cancer Center, Union City, CA. Retrieved from https://www.lexjansen.com/wuss/2003

Catalano, P., Chikritzhs, T., Stockwell,T., Webb, M., Rohlin C-J., & Dietze, P. (2001) . *Trends in per capita alcohol consumption in Australia 1990/91 to 1998/99. Bulletin No. 4.* Perth, Australia: National Drug Research Institute, Curtin University of Technology. Retrieved from https://ndri.curtin.edu.au

Celik, H.E. & Yılmaz, V.(2013). LISREL 9.1 *ile Yapsil Esitlik Modellemesi: Temel Kavramlar-Uygulamalari* (2. Baski). Ankara: Pegem Akademi

Chapter 3: Genetic and Psychosocial Influences (2000). In D.E. Shalala (Secretary U.S Department of Health and Human Services). Tenth *Special Report to the U.S. Congress on Alcohol and Health*. (pp. 159-191). Retrieved from https://pubs.niaaa.nih.gov

Chen, M-J., Grube, J.W., Bersamin, M., Waiters, E., & Keefe, D. B (2005). Alcohol Advertising: What Makes It Attractive to Youth? *Journal of Health Communication*, 10:553–565. DOI: 10.1080/10810730500228904

Chermack, S.T. & Giancola, H.M. (1997). The relation between alcohol and aggression: An integrated bio-psychosocial conceptualization. *Journal of Clinical Psychology Review 17*, 621–649.

Chowell, G., Hyman, J.M., Bettencourt, L.M.A., &Castillo-Chavez, C (Eds.). (2009). *Mathematical and Statistical Approaches in Epidemiology*. doi: 10.1016/j.phrp.2011.08.005

Cintron-Arias, A., Sanchez, F., Wang, X., & Castillo-Chavez, C. (2009). The role of Nonlinear relapse on Contagion Amongst Drinking Communities, In G. Chowell, L.M.A. Bettencourt, J.M. Hyman & C. Castillo-Chavez (Eds.), *Mathematical and Statistical Estimation Approaches in Epidemiology* (pp. 343-360). Springer Netherlands. DOI: 10.1007/978-90-481-2313-1_14

Civelek, M. E (2018). *Essentials of Structural Equation Modelling.* Downloaded from https://digitalcommons.unl.edu/zeabook/64

Cloninger, C.R., Sigvardsson, S. and Bohman, M.(1996). Type I and Type II alcoholism: An update. *Alcohol Health Res World 20*(1):18–23.

Collett, D. (1994). *Modelling Survival Data in Medical Research*. London: Chapman & Hall.

Cooper, M.L. (1994). Motivations for Alcohol Use among Adolescents: Development and Validation of a Four-Factor Model. *Psychological Assessment, 6* (2), *117-128. DOI: 10.1037/1040-3590.6.2.117.*

Cordeiro, G.M. and Lemonte, A.J. (2011). The Birnbaum-Saunders distribution: An improved distribution for fatigue life modelling. *Computational Statistics and Data Analysis, 55,* 1445-1461.

Cornah, D. (2006). *"Cheers" Understanding the relationship between alcohol and mental health.* London. The Mental Health Foundation. Downloaded from https://www.mentalhealth.org.uk

Corrao, G., Bagnardi, V., Zambon, A., & Arico, S. (1999). Exploring the dose–response relationship between alcohol consumption and the risk of several alcohol related conditions: a meta-analysis. *Journal of Addiction, 94,* 1551–1573.

Cox, R. B. (2007). *Toward an eco-developmental theory of adolescent substance use in Venezuela* (Doctoral Dissertation). Available from Dissertations & Theses database. (UMI No. 3282080).

Crawford, D. M., Whitbeck, L. B., Hoyt, D. R. (2009). Propensity for Violence Among Homeless and Runaway Adolescents: An Event History Analysis. *Crime & Delinquency.* https://doi.org/10.1177/0011128709335100

Crundall, I. (1995). Perceptions of alcohol by student drinkers at university. *Drug and Alcohol review*, 14(4), 363-368. DOI: 10.1080/09595239500185491

Danelia, K (2011). *Modelling early alcohol initiation: A comparison of linear regression, logistic regression, and discrete time hazard models* (Doctoral Dissertation Oklahoma State University). Downloaded from https://core.ac.uk

Dawson, D. A., Grant B.F., Stinson, F.S., Chou, P.S., Huang, B. & Ruan, W (2005). Recovery from DSM–IV Alcohol Dependence: United States, 2001–2002. *Addiction 100*(3), 281-292. DOI: 10.1111/j.1360-0443.2004.00964.x

Delgado-Lobete, L., Montes-Montes, R., Vila-Paz, A., Cruz-Valiño, J-M., Gándara-Gafo, B., Talavera-Valverde, M-A & Santos-del-Riego, S. (2020). Individual and Environmental Factors Associated with Tobacco Smoking, Alcohol Abuse and Illegal Drug Consumption in University Students: A Mediating Analysis, *International Journal of Environmental Research and Public Health, 17,* 3019; doi:10.3390/ijerph17093019

Desmond, A. (1985). Stochastic Models of failure in random environments. The Canadian Journal of Statistics, 13(2), 171-183. https://doi.org/10.2307/3315148

DeWit, D.J., Adlaf, E.M., Offord, D.R., &. Ogborne, A.C. (2000). Age at First Alcohol Use: A Risk Factor for the Development of Alcohol Disorders. *American Journal of Psychiatry, 157*, 745–750.

Dey, T., Mukherjee, A. & Chakraborty, S. (2020). A Practical Overview and Reporting Strategies for Statistical Analysis of Survival Studies: An Overview of Study Design and Statistical Considerations. Supplement *Chest journals* 158(1S):S39-S48 DOI: https://doi.org/10.1016/j.chest.2020.03.015

Donovan, J.E (2004). Adolescent Alcohol Initiation: *A Review of Psychosocial Risk Factors. Journal of Adolescent Health. 35,* 529.e7 - 529.e18

Donovan, J.E. (1977). *A typological Study of Self-Reported Deviance in a national sample of Adolescents* (Doctoral Dissertation). Available from University of Colorado, Dissertation Abstracts, 1977, 38, p.24426-B [05]).

Dubowski, K.M. (1985). Absorption, distribution and elimination of alcohol: highway safety aspects. *Journal of Studies on alcohol*, Supplement, (s10), 98-108. https:/doi.org/10.15288/jsas.1985.s10.98

Dupuis, D. J. and Mills, J. E. (1998). Robust estimation of the Birnbaum-Saunders distribution. *IEEE Transactions on Reliability, 47*, 88-95.

Egan, J.R & Hall, I.N. (2015). A review of back-calculation techniques and their potential to inform mitigation strategies with application to non-transmissible acute infectious diseases. *Journal of Royal Society Interface 12:*20150096. http://dx.doi.org/10.1098/rsif.2015.0096

Eze, J.I (2009). *Modelling HIV/AIDS Epidemic in Nigeria* (Doctoral Dissertation, University of Glasgow, Scotland). Retrieved https://theses.gla.ac.uk

Fergusson, D.M., Boden, J.M., & Horwood, L.J. (2009). Tests of Causal Links Between Alcohol Abuse or Dependence and Major Depression. *Archive of General Psychiatry. 66* (3), 260-266. Downloaded From: *https://jamanetwork.com/ on 07/31/2019*

Ferreira, M., Gomes, M.I. and Leiva V. (2012). On an Extreme Value Version of the Birnbaum-Saunders Distribution. *REVSTAT-Statistical Journal, 10*(2), 181-210. Retrieved from https://www.researchgate.net

Field, J.B .F. (1985). *A statistical study of the distribution of alcohol consumption and consequent inferential problems* (Doctoral thesis, University of Adelaide, Adelaide, Australia). Retrieved from https://digital.library.aldelaide.edu.au/

Fillmore, K.M. (1988). *Alcohol use across the life course: a critical review of 70 years of international longitudinal research.* Addiction Research Foundation, Toronto.

Finkelstein, M. (2008). *Failure rate modelling for reliability and risk.* Retrieved from http://www.springer.com/978-1-84800-985-1.

Fox, J. (2002). *Structural Equation Models. Appendix to An R and S-PLUS companion to Applied Regression.* Thousand Oaks, CA: Sage.

French, D.A., Teymuroglu, Z., Lewis, T.J., & Braun, R.J. (2010). An integro-Differential equation model for the spread of alcohol abuse. *Journal of integral equations and applications, 22* (3), *441-462*. Retrieved from https://projecteuclid.org/

Fukuwaka, M. (1996).Allometric Back-Calculation of Individual Growth for Chum Salmon Otolith During Early Life. S*cientific Reports of the Hokkiudo Salmon Hatchery No 50* 113-116,

Gail, M.H. & Brookmeyer, R (1989). Projecting the incidences of AIDS. *JAMA, 263*(11), 1538-1539. doi:10.1001/jama.1990.03440110104037

Gail, M.H. (2005). Risk. In P. Armitage. & T. Colton (Eds.), *Encyclopaedia of Biostatistics (*2nd ed.). New York, NY: John Wiley & Sons

Gao, B. & Bataller, R. (2011). Alcoholic liver disease: pathogenesis and new therapeutic targets. *Journal of Gastroenterol, 141*(5), 1572-85.

Gmel, G., & Rehm, J. (2003). Harmful alcohol use. *Journal of Alcohol Research and Health, 27*, 52–62

Gómez, A.P., Días-Granados, O.S. & Alarcón, L.F. (2011). Age at Onset of Alcohol Consumption and Risk of Problematic Alcohol and Psychoactive Substance Use in Adulthood in the General Population in Colombia. *The Journal of International Drug, Alcohol and Tobacco Research. 1* (1), 19–24.

Gordis, E. (1997). "The Etiology, Consequences, and Treatment of Alcoholism". National Institute on Alcohol Abuse and Alcoholism. *Journal of Liver Transplantation and Surgery, 3*(3), 199-205.

Gowin, J.L., Sloan, M.E., Stangl, B.L., Vatsalya, V. & Ramchandani, V.A. (2017). Vulnerability for Alcohol Use Disorder and Rate of Alcohol Consumption. *Am J Psychiatry 174* (11); 1094-1101.

Grant, B.F. & Dawson D.A (1997). Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiological Survey. *Journal of substance abuse 9*, 103-110.

Green, C.A., Polen, M.R.,& Perrin, N.A. (2009). Structural Models of Gender, Alcohol Consumption, and Health. *Journal of Substance Use & Misuse.* https://doi.org/10.1081/JA-120016568

Guiraud, P., Leiva, V. & Fierro, R. (2009). A non-central version of the Birnbaum-Saunders distribution for reliability analysis. *IEEE Transaction on Reliability, 58*, 152–160.

Hassan, M.N (2013*). Factors Associated With Alcohol Abuse among University of Nairobi Students* (Master's thesis, University of Nairobi, Kenya). Available in https://erepository.uonbi.ac.ke. Retrieved from https://pdfs.semanticscholar.org

Hingson, R.W., Heeren, T. & Winter, M.R. (2006). Age at Drinking Onset and Alcohol Dependence. Age at Onset, Duration, and Severity. *Archive of Paediatric and Adolescence Medicine. 160:* 739-746

Hoffmann, J.P. & Cerbone, F.G. (2002). Parental substance use disorder and the risk of adolescent drug abuse: an event history analysis. *Drug and Alcohol Dependency, 66(3)*, 255-264.

Holahan, C.J., Schutte, K.K., Brennan, P.L., Holahan, C.K., Moos, B.S. & Moos R.H (2010).Late-Life Alcohol Consumption and 20-Year Mortality. *Alcoholism: Clinical and Experimental Research, 34* (11): 1-11. DOI: 10.1111/j.1530-0277.2010.01286.x

Holtby, L.B., Anderson, B.C., & Kadowaki, R.K. (1990). Importance of smolt size and early ocean growth to international variability in marine survival of coho salmon (Oncorhynchus kisutch). *Canadian Journal of Fisheries and Aquatic Science, 47*, 1281-2194.

Horihan, A. (2014). Gender Differences and Impacts on Addiction Treatment: Special Considerations for Working with Female Clients. Counselor Education Capstones. 10. Master Thesis, Winona State University, Minnesota, USA. Downloaded from https://openriver.winona.edu/counseloreducationcapstones/10

Howie, E.K., Sui, X., Lee, D., Hooker, S.P. Hébert, J.R. & Blair S.N. (2011). Alcohol Consumption and Risk of All-Cause and Cardiovascular Disease Mortality in Men. *Journal of Aging Research Volume 2011*, Article ID 805062, doi:10.4061/2011/805062.

Hoyle, R. (1995). The structural equation modelling approach: Basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural Equation Modelling: Concepts, issues and applications* (pp. 1-15). Thousand Oaks, CA: Sage.

Humphreys, J.H., Warner, A., Costello, R., Lunt, M., Verstappen, S.M.M. &Dixon, W.G (2017). Quantifying the hepatoxic risk of alcohol consumption in patients with rheumatoid arthritis taking methotrexate. *Annals of the Rheumatic Diseases, 76*(9). https://dx.doi.org/10.1136/annrheumdis-2016-210629

ICAP (2009) International Center for Alcohol Policies. *Determinants of Drinking.* Accessed from www.icap.org

Isham, V (1989). Estimation of incidence of HIV infection. *Philosophical Transactions of Royal Society of London B* (325), 113-121

Iwanickal, K. A. & Olajossy, M. (2015). The concept of alcohol craving [Electronic version]. *Journal of Psychiatry Pol., 49*(2), 295–304. Retrieved from http://www.psychiatriapolska.pl

Jayaram, J., Kannan, V. R., & Tan, K.C (2004). Influence of Initiators on Supply Chain value Creation. *International Journal of Production Research, 42*(20), 4377-4399 Retrieved from https://www.researchgate.net DOI: 10.1080/00207540410001716516

Jellinek, E. M. (1960). The Disease Concept of Alcoholism. New Haven: Hill house Press. Retrieved from http://books.google.com/books (Original work published 1960,)

Jernigan, D. (n.d). A Global Perspective On Drinking Ages And Alcohol Use. Retrieved from www.dartmouth.edu/

Jewell, N.P. (1990). Some statistical issues in studies of epidemiology of AIDS. *Journal of Statistics in Medicine 9,* 1387-1416.

Johnson, N., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2. New York: Wiley.

Kaithuru, P. N., & Stephen, A. (2015). Alcoholism and its Impact on Work Force: A Case of Kenya Meteorological Station, Nairobi. *Journal of Alcohol Drug Dependence*. 3: 192. doi: 10.4172/2329-6488.1000192.

Katikireddi, S.V., Whitley, E., Lewsey, J., Gray, L. & Leyland, A.H.(2017). *Socioeconomic status as an effect modifier of alcohol consumption and harm: analysis of linked cohort data*. Retrieved from www.thelancet.com/public-health ,2:e267- e276 . http://dx.doi.org/10.1016/ S2468-2667(17)30078-6

Keller, P. (2016). *Alcohol: Does it make you successful? A longitudinal analysis* (Master Thesis, University of Zurich). SOEP — The German Socio-Economic Panel study at DIW Berlin. Retrieved from https://www.diw.de

Kenya Gazette Supplement Senate Bills, (2014). Special Issue. NAIROBI, 28th February, 2014 *Kenya Gazette Supplement No. 17 (Bills No. 5)*. The Alcoholic drinks Control (Amendment) Bill, 2014. Retrieved from www.kenyalaw.org

Kessels, R. & Erreygers, G (2016). Structural equation modeling for decomposing rank-dependent indicators of socioeconomic inequality of health: an empirical study. *Health Economics Review*, 6:56, DOI 10.1186/s13561-016-0134-2

Khine, M.S. (ed.), (2013). *Application of Structural Equation Modelling in* Educational Research and Practice. Rotterdam, Sense publishers. Retrieved from https://www.researchgate.net/publication

King, A.J (2014). Bayesian Event History Analysis with Applications to Recurrent Episodes of Illicit Drug Use. University of California, Los Angeles USA

Kline, R.B. (2011). *Principles and Practice of Structural Equation Modelling*, 3rd Ed. Methodology in the Social Sciences series. New York: The Guilford Press.

Koning, S.H., Gansevoort, R.T., Mukamal, K.J., Rimm, E.B., Bakker, S.J.L., Joosten, M.M., & PREVEND Study Group (2015). Alcohol consumption is inversely associated with the risk of developing chronic kidney disease. *Kidney International 87*, 1009–1016.

Korhonen, M. (2005). *Alcohol Problems and Approaches: Theories, Evidence and Northern Practice National Aboriginal Health Organization*. Retrieved from https://www.ruor.outtawa.ac/

Kraemer, H.C., Kazdin, A.E., Offord, D.R., et al. (1997). Coming to terms with terms of risk. *Archive of General Psychiatry; 54*, 337–43. doi:10.1001/arcpsyc.1997.018301600065009

Kucera, O. & Cervinkova, Z. (2014). Experimental models of non-alcoholic fatty liver disease in rats. *World Journal of Gastroenterol, 20*(26), 8364-8376. Retrieved from http://www.wjgnet.com/1007-9327/full/v20/i26/8364.htm

Kuhn, S., Mascharek, A., Banaschewski, T., Bodke, A., Bromberg, U., Bu¨ chel, C., … , IMAGEN Consortium (2019). Predicting behaviour development of adolescent drinking from whole brain structure at 14 years of age DOI: https://doi.org/10.7554/eLife.44056

Kundu, D., Kannan, N., Balakrishnan, N. (2008). On the hazard function of Birnbaum-Saunders distribution and associated inference. *Computational Statistics and Data Analysis, 52,* 2692-2702.

Lecture 1(n.d) Introduction to Multilevel Models. *Lecture notes*. Course web site http://www.biostat.jhsph.edu/~fdominic/teaching/bio656/ml.html

Leiva, V. (2016). *The Birnbaum-Saunders Distribution.* London: Elsevier.

Leiva, V., Barros, M., Paula, G. & Sanhueza, D. (2008). Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics, 19,* 235–249.

Leiva, V., Lillo, C., Gomes, I.M., & Ferreira, M. (2018), Discussion of Birnbaum-Saunders distribution: a review of models, analysis, and applications and a novel financial extreme value data analytics from natural disasters. *Applied stochastic models in business and industry.* https://doi.org/10.1002/asmb.2400.

Leiva, V., Ponce, M., Marchant, C., & Bustos, O. (2012). Fatigue statistical distributions useful for modelling diameter and mortality of trees. *Colombian Journal of Statistics. 35,* 349–367.

Leiva, V., Sanhueza, A. & Angulo, J. (2009). A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stochastic Environmental Research and Risk Assessment, 23*, 299–307.

Lemonte, A.J. (2016). A note on the Fisher Information matrix of the Birnbaum-Saunders distribution. *Journal of Statistical Theory and Applications,* 15, 196-205. Downloaded from https://www.atlantis-press.com/journals/jsta/25856828

Li, C., DiPiro, N.D. and Krause, J (2017). A latent structural equation model of risk behaviors and pressure ulcer outcomes among people with spinal cord injury. *Spinal Cord*, 55, 553–558

Lilleskov, M.A., & Chakua, C (2013). Assessment Of Alcohol And Drug Use In The Private Sector In Kenya. *NACADA.* Retrieved from www.nacada.go.ke

Lim, S.S., Vos, T., Flaxman, A.D., Danaei, G., Shibuya, K., Adair-Rohani, H.,…, Ezzati, M (2012). A comparative risk assessment of burden of disease and injury attributable to risk factors and risk factor clusters in 21 regions, 1990- 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet. 380* (9859), 2224-60. DOI: 10.106/S0140-6736(12)61766-8

Loring, B (2014). Alcohol and inequities: Guidance for addressing inequities in alcohol-related harm. *World Health Organization.* Downloaded from https://www.euro.who.int

Ma, C & Brunt, E.M. (2012). Histopathologic evaluation of liver biopsy for cirrhosis. *Journal of Advanced Anatomy and Pathology, 19*(4), 220-30.

MacCallum, R.C. & Austin, J.T. (2000). Applications of structural equation modelling in psychological research. *Annual Review of Psychology, 51*(1), 201-226. https://doi.org/10.1146/annurev.psych.51.1.201

Maggs, J.L., Staff, J., Patrick, M.E. & Wray-Lake, L. (2019). Very early drinking: Event history models predicting alcohol use initiation from age 4 to 11 years. *Addictive Behaviors, 89,* 121-127, https://doi.org/10.1016/j.addbeh.2018.09.030.

Mahugu (n.d). *The factors contributing to alcohol abuse and alcoholism* (Master's thesis*).* Moi University, Eldoret. Kenya. Downloaded www.kapc.or.ke

Maithya, R., Okinda, R., & Mung'atu, F. (2015). A Baseline Survey on Effects of Drug and Substance Abuse on Academic Participation Among Students In Technical Institutions In Kenya. *International Journal of Education and Research. 3*: 629-642.

Maitso, S.A., Clifford, P.R., Stout, R. L. & Davis, C.M. (2008). Factors Mediating the Association between Drinking in the First Year After Alcohol Treatment and Drinking at Three Years. *Journal of Study of Alcohol and Drugs 69*: 728-737.

Malacane, L. A., Murphy, S. A., & Collins, L. M. (1997). Determining the risk set for hazard functions when survival methods are applied to social science and prevention data. University Park, PA: The Methodology Center, Pennsylvania State University.

Mann, K., Hermann, D., & Heinz, A. (2000). One Hundred Years Of Alcoholism: The Twentieth Century. *Journal of Alcohol & Alcoholism, 35*(1), 10–15. Retrieved from https://www.ncbi.nlm.nih.gov/

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. New York: John Wiley & sons. Retrieved from http://books.google.com/books (Original work published 1974)

Manthey, J.L., Aidoo, A.Y. & Ward, K.Y. (2008). Campus drinking: an epidemiological model. *Journal of Biological Dynamics, 2*(3), 346-356. Retrieved from http://www.tandfonline.com/loi/tjbd20

Marchant, C (2016). *Essays on multivariate generalised Birnbaum-Saunders methods* (Doctoral thesis, Universidade Federal de Pernambuco, Chile). Retrieved from https://researchgate.com

Marchant, C., Leiva, V., Athayde, E.,& Azevedo, C. (2011). Modelling wind energy flux by a Birnbaum-Saunders distribution with unknown shift parameter. *Journal of Applied Statistics 38,* 2819–2838.

McKee, M., Shkolnikov, V., & Leon, D.A. (2001). Alcohol is implicated in the fluctuations in cardiovascular disease in Russia since the 1980s. *Annals of Epidemiology, 11*, 1–6.

Merchant, F.Z. (2013).*A study on the depiction of drug usage, alcohol consumption and cigarette smoking in movies and its perceived effect on a young audience. A comparative study of American and Indian cinema and their respective Audiences* (Doctoral Thesis, University of South Florida: USA). Retrieved from http://scholarcommons.usf.edu/etd/4826

Merline, A., Jager, J & Schulenberg, J.E (2008). Adolescent risk factors for adult alcohol use and abuse: stability and change of predictive value across early and middle adulthood. *Journal Society for the Study of Addiction. Addiction, 103 (Suppl. 1),* 84–99

Merriam-Webster(n.d.) retrieved from https://www.merriam-webster.com/dictionary/disorder

Meydan, C.H. & Şen, H. (2011). *Structural Equation Modelling AMOS Applications*. Ankara: Detail Publishing.

Midgette, G & Kilmer, B (2015). *The Effect of Montana's 24/7 Sobriety Program on DUI Re-arrest: Insights from a Natural Experiment with Limited Administrative Data*. Working Paper RAND Justice, Infrastructure, and Environment. WR-1083-MHP.

Miner, M. A. (1945). Cumulative damage in fatigue. *Journal of Applied Mechanics 12*, 159–164.

Moallem, N. R., Courtney, K. E., Bacio, G. A. & Ray, L.A (2013). Modeling alcohol use disorder severity: an integrative structural equation modeling approach, *Frontiers in Psychiatry/ | Addictive Disorders and Behavioral Dyscontrol.* doi:10.3389/fpsyt.2013.00075

Moffitt, T.E. (1993). Adolescence-limited and life-course-persistent antisocial behaviour: a developmental taxonomy. *Psychology review, 100*(4), 674-701. .DOI: 10.1037/0033-295X.100.4.674

Mordey, A. (2015). What Are The Risk Factors For Alcohol And Drug Addictions? Retrieved from https://www.thecabinchiangmai.com

Morse, R.M., & Flavian, D.K (1992) The definition of alcoholism. The Joint Committee of the National Council on Alcoholism and Drug Dependence and the American Society of Addiction Medicine to Study Definition and Criteria for the Diagnosis of Alcoholism. *JAMA*, *268*, 1012-1014. doi:10.1001/jama.1992.03490080086030

Mou, Y., & Lin, C.A. (2020) Consuming Alcohol to Prepare for Adulthood: An Event History Analysis of the Onset of Alcohol Use Among Chinese College Students. *Sage, July-Sept, 1-11*. https://doi.org/10.1177/2158244020939911

Mumenthaler, M.S., Taylor, J.L., O'Hara, R. and Yesavage, J.A. (1999). Gender Differences in Moderate Drinking Effects. *Alcohol Research & Health 23*(1), 55-64

Murray, R.P., Connett, J.E., Tyas, S.L. Bond, R., Ekuma, O., Silversides, C.K. & Barnes, G.E (2002). Alcohol volume, drinking pattern, and cardiovascular disease morbidity and mortality: is there a U-shaped function? *American Journal of Epidemiology 155*(3), 242–248. DOI:10.1093/aje/155.3.242

NACADA (2010). Alcohol Use in Central Province of Kenya: A Baseline Survey On Magnitude, Causes And Effects From The Perspective Of Community Members And Individual Users.

NACADA, (2012). Report on Rapid Situation Assessment of drug and substance abuse in Kenya. Nairobi. Government Printer

NACADA (n.d.). *National Survey On Alcohol And Drug Abuse Among Secondary School Students In Kenya*. Retrieved from www.nacada.go.ke

Nacoa (n.d.) "The National Association for Children of Alcoholics". *Children of alcoholics important facts.* Downloaded from http://www.nacoa.net/impacts.htm

Neels, K (2014). Event-history analysis: discrete- & continuous-time methods. Society for Longitudinal & Life-course Studies. Summer School, University of Amsterdam, 25-29 August 2014, Sociology Department, University of Antwerp QASS-Programme, KULeuven. Retrieved from https://fsw.vu.nl

Nees,F., Tzschoppe, J., Patrick, C.J., Vollsta'dt-Klein, S., Steiner, S., Poustka, L., … the IMAGEN Consortium (2012). Determinants of Early Alcohol Use In Healthy Adolescents: The Differential Contribution of Neuroimaging and Psychological Factors. *Neuropsychopharmacology (2012) 37*, 986–995. doi:10.1038/npp.2011.282.

Ng, H., Kundu, D., Balakrishnan, N., 2003. Modified moment estimation for the two-parameter Birnbaum–Saunders distribution. *Computational. Statistics. Data Analysis. 43*, 283–298.

Nicholson, A., Bobak, M., Murphy, M., Rose, R. & Marmot, M (2005). Alcohol consumption and increased mortality in Russian men and women: a cohort study based on the mortality of relatives. *Bulletin of the World Health Organisation; 83*(11): 812-819

O'shea,R.S., Darasathy,S. & Mccullough, A.J. (2010). Alcoholic Liver Disease and the Practice Guideline Committee of the American Association for the Study of Liver Diseases: The Practice Parameters Committee of the American College of Gastroenterology. *Journal of Hepatology,* 307-328.

Odongo D, (2020, 24–30 April). Kenya a drinking Nation. *The Nairobian Newspaper,* p. 4+, www.standardmedia.co.ke

Pal, P and Ray, S. (2016). Alcoholic Liver Disease: A Comprehensive Review. *European Medical Journal, 1*(2), 85-92.

Palmer, T.N. (Ed.). (1991). *Alcoholism: A molecular perspective*. New York: Plenum Press

Panda, U.N. (Ed) (2009). *Concise Pocket Medical Dictionary* 2nd ed. Jaypee Brothers Medical Publishers (P) Ltd. New Delhi, India

Petersen, T. (1995). "*Analysis of Event Histories". Handbook of Statistical Modelling for the Social and Behavioural Sciences*. New York: Plenum Press.

Pettigrew, S & Donovan, R. (2003). *A literature Review of the factors that influence Alcohol consumption and the effectiveness of the past interventions*. Downloaded from www.academia.edu

Podlaski, R. (2008). Characterization of diameter distribution data in near natural forests using the Birnbaum-Saunders distribution. *Canadian Journal of Forest Research 18*, 518–527.

Pollard, M., Green, H. D., Kennedy, D.P., Go, M and Tucker, J.S (2013). *Adolescent Friendship Networks and Trajectories of Binge Drinking.* Population Research Centre. DOI:10.2139/ssrn.2286295 downloaded from https://www.researchgate.net

Ramsden, E .(2015). Making animals alcoholic: shifting laboratory models of addiction. *Journal of the History of the Behavioural Sciences,* 51(2), 169-94. Retrieved from https://www.ncbi.nlm.nih.gov/

Ravanan, R. & Venkatesan, P.(2008). Some new approaches for modelling the incubation period of HIV/AIDS epidemic. International *Journal of Computer, Mathematical Sciences and Applications. Serials Publications, 2(*4), 223-237.

Raykov, T. & Marcoulides, G.A. (2006). *A First Course in Structural Equation Modelling.* 2nd edition, London: Lawrence Erlbaum Associates.

Read, J.P., Wood, M.D., Kahler, C.W., & Maddock, J.E. (2003). Examining the role of drinking motives in college student alcohol use and problems. *Psychology of Addictive Behaviours, 17*(3), 13-23. Downloaded from www.googlescholar.com

Rehm, J. (1998b). Measuring alcohol consumption: how about adopting usual epidemiological standards. *Journal of Addiction 93,* 970–972.

Rehm, J., Baliunas,D., Borges, GLG., Graham, K., Irving, H., Parry, C.D., … Poznyak, V. (2010a). The relation between dimensions of alcohol consumption and burden of disease: an overview. *Addiction, 105* (5), 817-843

Rehm, J., Gmel, G., Room, R., & Frick, U. (2001c). Average volume of alcohol Consumption, Drinking patterns and related burden of mortality in young people in established economies of Europe. *European Addiction Research, 7*(3), 148-151. DOI: 10.11159/000050732

Rehm, J., Kehoe, T., Gmel, G., Stinson, F., Grant, B., & Gmel, G (2010). Statistical modelling of volume of alcohol exposure for epidemiological studies of population health: the US example. *Journal of Population Health Metrics, 8*, 3. Downloaded from https://scholargoogle.com

Rehm, J., Patra, J., & Popova, S. (2006). Alcohol attributable mortality and potential years of life lost in Canada 2001: Implication for prevention and policy. *Addiction, 101,* 373-384

Rehm, J., Room, R., Graham, K., Monteiro, M., Gmel, G., & Sempos, C.T. (2003). The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease - an overview. *Journal of Addiction 98,* 1209–1228

Rehm, J., Sempos, C.T. (1995). Alcohol consumption and mortality—questions about causality, confounding and methodology. *Journal of Addiction 90*, 493–498.

Rehm, J., Sempos, C.T., & Trevisan, M. (2003a). Alcohol and cardiovascular disease—more than one paradox to consider. Average volume of alcohol consumption, patterns of drinking and risk of coronary heart disease: a review. *Journal of Cardiovascular Risk 10*, 15–20.

Rehm, J., Shield, K. D., Rehm, M. X., Gmel, G. & Frick, U. (2012). *Alcohol consumption, alcohol dependence and attributable burden of disease in Europe: Potential gains from effective interventions for alcohol dependence*. Canada, Centre for Addiction and Mental Health. Retrieved from www.camh.net

Rehm, J., Taylor, B., Roerecke, M., & Patra, J. (2007). Alcohol consumption and alcohol-attributable burden of disease in Switzerland, 2002. *International journal of Public Health, 52(6),* 383-392. DOI: https://doi.org/10.1007/s00038-007-7010-0

Rigdon, E. E. (1998). Structural Equation Modeling. In G. A. Marcoulides (Ed.), Modern Methods for Business Research (pp. 251-294). Mahwah, NJ: Erlbaum.

Rinne, H. (2014). *The hazard rate: Theory and inference (with supplementary MATLAB- programs).* Justus-Liebig-Universitt. Retrieved from https://core.ac.uk

Roerecke, M. & Rehm, J. (2012). Alcohol intake revisited: Risks and Benefits. *Current Atherosclerosis Reports, 14(6),* 556-562. DOI: 10.1007/s11883-012-0277-5

Rosenberg, P.S. & Gail, M.H. (1991). Back-calculation of Flexible Linear Models of the HIV Infection Curve. *Applied Statistics*, 40, 269-282

Rosenberg, P.S., & Gail, M.H. (1990). Uncertainty in estimates of HIV prevalence derived back-calculation. *Annals of Epidemiology*, 1:105-115

Rozin, L. & Zagonel, I.P.S. (2012). Risk factors for alcohol dependence in adolescents. *Journal of Acta Paul Enferm. 25(2):*314-8. DOI: 10.1590.S0103-21002012000200025

Rubin, R. (1999). Effect of ethanol on platelet function. *Alcoholism, Clinical and Experimental Research 23*, 1114–1118.

Rundel C. (2012). *Lecture 14: Hazard*. Statistics 104. Duke Statistical Science. Downloaded from https://www2.sta.duke.edu

Sampson, R. J., & Laub, J. H. (2003). Life-course disasters? Trajectories of crime among delinquent boys followed to age 70. *Criminology, 41*(3), 555-592.

Sánchez, F. (2007). *Studies In Epidemiology And Social Dynamics*. (Doctoral Dissertation, Cornell University). Retrieved from https://ecommons.cornell.edu/bitstream/handle/1813/34124/Sanchez.pdf?sequence

Sánchez, F., Wang, X., Castillo-Chávez, C., Gorman, D. M., & Gruenewald, P. J. (2007). Drinking as an epidemic--A simple mathematical model with recovery and relapse. In K. A. Witkiewitz & G. A. Marlatt (Eds.), *Therapist's guide to evidence-based relapse prevention* (pp. 353–368). Elsevier Academic Press. https://doi.org/10.1016/B978-012369429-4/50046-X

Sanhueza, A., Leiva, V. and Balakrishnan, N. (2008). The generalized Birnbaum-Saunders distribution and its theory, methodology and application. *Communications in Statistics: Theory and Methods, 37,* 645-670.

Sanhueza, A., Marchant, C., Leiva, V. & Cavieres, M. (2013). Air contaminant statistical distributions with application to PM10 in Santiago, Chile. *Reviews of Environmental Contamination and Toxicology 223*, 1–31.

Sarhan, A.M. & Kundu, D. (2009). Generalized Linear Failure Rate Distribution. *Communications in Statistics- Theory and Methods, 38* (5), 642-660. https://doi.org/10.1080/03610920802272414

Savalei, V and Bentler, P. M. (n.d). *Chapter 17.* In *Structural Equation Modelling*. University of California, Los Angeles. Retrieved from https://citeseerx.ist.psu.edu

Schaaf, D. & Scragg, R. (n.d). Alcohol consumption and associated risk factors in Auckland Pacific Island adolescents. Population health, University of Auckland. New Zealand. Downloaded from https://www.hpa.org.nz .

Schoenbach, V.J. & Rosamond, W.D. (2000).2000 ed. *Understanding the Fundamentals of Epidemiology: an evolving text*. Downloaded from www.epidemiolog.net

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*, 323–337.

Schuckit, M.A. (2002). Vulnerability factors for Alcoholism. In : Davis, K., Charney, D., Coyle, J.T., & Nemeroff, C.B. (Eds.) *Neuropsycholopharmacology: The Tifth Generation of Progress*, Philadelphia, PA: Lippincott Williams & Wilkins, 1399-1411

Sharma, S and Samanta, G.P. (2013). *Drinking as an epidemic: a mathematical model with dynamic behaviour*. Retrieved from https://www.researchgate.net/publication/264145264.

Sher, K.J., Martin, E.D., Wood, P.K., & Rutledge, P.C. (1997). Alcohol use disorders and neuropsychological functioning in first year undergraduates. *Experimental and Clinical Psychopharmacology, 5*(3), 304-315. https://doi.org/10.1037/1064-1297.5.3.304

Shirley, K.E., & et al. (2010). Hidden Markov Models For Alcoholism Treatment Trial Data. *The Annals of Applied Statistics, 4*(1), 366–395.

Shuval, K., Finley, C.E., Chartier, K.G., Balasubramanian, B.A., Gabriel, K.P. & Barlow, C.E. (2012). Cardiorespiratory Fitness, Alcohol Intake, and Metabolic Syndrome Incidence in Men. *Med Science Sports Exercise; 44*(11): 2125–2131. doi:10.1249/MSS.0b013e3182640c4e

Simwa, R.O., & Pokhariyal, G.P(2003). On empirical Modelling of HIV/AIDS Pandemic with Application to East Africa. *African Journal of Science and Technology (AJST). Science and Engineering Series, 4*(1), 104-109.

Singer, J. D., and Willett, J. B. (2003) *Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence.* https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Sjölund, S. (2015). *Cognitive Ability, Alcohol Use and Alcohol-Related Harm* (Doctoral Thesis, Karolinska Institutet, Stockholm, Sweden). Retrieved from https://openarchive.ki.se

Skog, O.J. (1985). The collectivity of drinking cultures: A theory of the distribution of alcohol consumption. *British Journal of Addiction 80,* 83–99. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3856453

Smith, G. D., Shaw,L.J., Maini, P.K., Ward, R.J., Peters, T.J., & .Murray, J.D.(1993). Mathematical modelling of ethanol metabolism in normal subjects and chronic alcohol misusers. *Journal of Alcohol & Alcoholism, 28*(1), 25-32.

Sprunt, H. (2004). *Blood Alcohol Concentration: Modelling with mathematics*. Downloaded from www.appstate.edu on 17-01-2019

Stamm, K. E (2007). *Structural Equation Modeling of Alcohol Issues in an Emergency Department Sample*. (Master Thesis, University of Rhode Island, Digital Commons @URI), Open Access Master's Theses. Paper 1589. https://digitalcommons.uri.edu/theses/1589

Staveren, W.A., Knoops, K.T.B., Groot de, L.C., Albert-Fidanza, A. & Kromhout, D (2006). Comparison of three dietary scores in relation to 10-year mortality in elderly European subjects: the HALE project. *European Journal of Clinical Nutrition, 60*: 746-755. DOI: 10.1038/sj.ejcn.1602378

Steele, F (2005). *Event History Analysis: A National Centre for Research Methods Briefing Paper*. Centre for Multilevel Modelling, Graduate School of Education, University of Bristol. Retrieved from https://www.eprints.ncrrm.ac.uk

Stewart, C. H. (2010). *Multilevel modelling of event history data: comparing methods appropriate for large datasets (*Doctoral dissertation, University of Glasgow, Scotland). Retrieved from http://theses.gla.ac.uk/2007/

Stowell, A.R. & Stowell, L.I. (1998). Estimation of Blood Alcohol Concentrations after Social Drinking. *Journal of Forensic Sciences, 43*(1), 14-21.

Suhr, (2006). *The Basics of Structural Equation Modelling*. Office of Budget and Institutional Analysis, University of Northern Colorado. Downloaded from https://www.lexjansen.com

Sutherland. A. (2012). Is parental socio-economic status related to the initiation of substance abuse by young people in an English city? An event history analysis. *Social Science & Medicine, 74 (7),* 1053-1061 https://doi.org/10.1016/j.socscimed.2011.12.026

Sznitman, S. & Engel-Yeger, B (2017). Sensation Seeking and Adolescent Alcohol Use: Exploring the Mediating Role of Unstructured Socializing With Peers. *Alcohol and Alcoholism,* 52(3) 396–401, doi: 10.1093/alcalc/agx008

Tekle, F. B., & Vermunt, J. K. (2012). Event history analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 3. Data analysis and research publication* (p. 267–290). American Psychological Association. https://doi.org/10.1037/13621-013

Teymuroglu, Z. (2009). Continuum models for the spread of alcoholism. *Journal of integral equations and applications*. Retrieved from https://www.math.ucdavis.edu/

Tripathy, J. P., Thakur, J. S., Jeet, G., & Jain. S. (2018). Structural equation modeling to identify the risk factors of diabetes in the adult population of North India*Tropical Medicine and Health* 46:23 https://doi.org/10.1186/s41182-018-0104-y

Tuma, N.B. & Hannan, M.T. (1979). Methods for Temporal analysis. *Annual review of sociology, 5*(1), 303-328. Retrieved from https://.jstor.org/stable/2945957

Tuyns, A. J., Pequignot, G., & Jensen, O. M. (1977). [Esophageal cancer in Ille-et-Vilaine in relation to levels of alcohol and tobacco consumption. Risks are multiplying]. *Bulletin du Cancer, 64*(1), 45-60

UNAIDS (2010). *Methods for estimating HIV incidence, Quarterly update on HIV epidemiology*. Geneva: UNAIDS

Vaillant, G.E. (2003). A 60-year follow-up of alcoholic men. *Journal of Addiction, 98,* 1043–1051.

Vaillant, G.E., & Hiller-Sturmhofel, S. (1996). Natural history of alcoholism. *Alcohol Health and Research World, 20,* 152–161.

Venkatesan, P. (2006). A Comprehensive Back-calculation Framework for Estimation and Prediction of HIV/AIDS in India. *Journal of Communicable Diseases, 38* (1), 40-56.

Wagner, F.A., Velasco-Mondragon, H.E., Herrera-Vazquez, M, et al., (2005). Early alcohol or tobacco onset and transition to other drug use among students in the state of Morelos, Mexico. *Drug alcohol dependence, 77,* 93-96.

Walters, C.E., Straughan, B. and Kendal, J.R. (2013). 'Modelling alcohol problems: total recovery'. Journal of Ricerche di matematica, 62 (1), 33-53. Retrieved from http://dro.dur.ac.uk

Walter H, Dvorak A, Gutierrez K, Zitterl W, Lesch OM. (2005). Gender Differences: Does Alcohol Affect Females More Than Males? Neuropsychopharmacologia Hungarica. VII/2: 78-82.

Wambui, M. (2015, August 25). National Authority for the Campaign Against Drug Abuse: We have 2.65 million drug addicts in Kenya, *Standard Digital*. Retrieved from www.standardmedia.co.ke.

Ward, B.R., Slaney, P.A., Facchin, A.R., & Land, R.W. (1989). Size-Biased Survival in Steelhead Trout (Oncorhynchus mykiss): Back-calculated Lengths from Adults' Scales Compared to Migrating Smolts at the Keogh River, British Columbia. *Journal of Fisheries and Aquatics Sciences, 46*(11), 1853-1858. https://doi.org/10.1139/f89-233

Washbrook, E. & Steele, F (2013). Discrete-time Event History analysis LECTURES. University of Bristol. Downloaded from www.bristol.ac.uk

White, H.R. & Jackson, K. (2004/2005). Social and psychological influences on emerging adult drinking behaviour. Alcohol research & Health, 28, 182-190. Downloaded from www.googlescholar.com

White, W. & Gasperin, D. (2006). Addiction, treatment and recovery: A primer for criminal justice personnel. Springfield, IL: University of Illinois-Springfield Center for legal and Policy Studies. Retrieved from www.williamwhitepapers.com

Whitman, I.R., Agarwal, V., Nah, G., Dukes, J.W., Vittinghoff, E., Dewland, T.A. & Marcus, G.M. (2017) Alcohol Abuse and Cardiac Disease. *Journal of the American College of Cardiology:* http://dx.doi.org/10.1016/j.jacc.2016.10.048

WHO Regional Office for Europe (2009). *Evidence for the effectiveness and cost–effectiveness of interventions to reduce alcohol-related harm*. Copenhagen, WHO Regional Office for Europe. Retrieved from http://www.euro.who.int/_data/assets/pdf_file/0020/43319/E92823.

WHO. (1999). Global status report on alcohol. World Health Organization, Substance Abuse Department, Geneva. Retrieved from https://scholargoogle.com

WHO. (2000). International guide for monitoring alcohol consumption and related harm. World Health Organization, Department of Mental Health and Substance Dependence. Copenhagen. Retrieved from https://scholargoogle.com

WHO. (2001). Surveys of drinking patterns and problems in seven developing countries. World Health Organization, Department of Mental Health and Substance Dependence. Geneva. Retrieved from https://scholargoogle.com

WHO. (2002). World Health Report 2002: reducing risks, promoting healthy life. World Health Organization, Geneva. Retrieved from https://scholargoogle.com

WHO. (2005). Alcohol, Gender and Drinking Problems: Perspectives from Low and Middle Income Countries. Department of Mental Health and Substance Abuse. Geneva. Retrieved from https://scholargoogle.com

WHO. (2011). Global status report on alcohol and health. Geneva: WHO Press. Retrieved from http://www.who.int/globalatlas/alcohol

WHO. (2014). *Global status report on alcohol and health* – 2014 ed. Geneva: WHO. Retrieved from https://scholargoogle.com

WHO. (2016). International statistical classification of diseases and related health problems. Fifth edition, 2016. Retrieved from https://scholargoogle.com

Wilfred de Alwis, N.M. & Day, C.P.(2007). Genetics of alcoholic liver disease and non-alcoholic fatty liver disease. *Journal of Semin Liver Diseases. 27*(1), 44-54.

Willett, J.B. and Singer, J.D. (1993). Investigating Onset, Cessation, Relapse, and Recovery - Why You Should, and How You Can, Use Discrete-Time Survival Analysis to Examine Event Occurrence. *Journal of Consulting and Clinical Psychology, 61*(6), 952-965.

Wilson, M., Stearne, A., & Gray, D (2010). The harmful use of alcohol amongst Indigenous Australian. *Australian Indigenous HealthInfoNet*. Retrieved from https://espace.curtin.edu.au

Wilson, R.A., Buchanan, J.R., Gleeson, J., & Braun, R.(2004). *A network model of alcoholism and alcohol policy*. School of Urban Affairs and Public Policy, University of Delaware. Proc. Math. Problems in Industry workshop, Nov 5 2004. Retrieved from www.eaton.math.rpi.edu

Woo, G.A & O'Brien, C. (2012). Long-term management of alcoholic liver disease. *Journal of Clinical and Liver Disease, 16* (4), 763-81.

Xie, H., McHugo, G., Drake, R., Sengupta, A. (2003) Using Discrete-Time Survival Analysis to Examine Patterns of Remission from Substance Use Disorder Among Persons with Severe Mental Illness. *Mental Health Services Research,* 5, pages55–64

Xu, R (2004). Proportional Hazards Mixed Models: A Review with Applications to Twin Models. *Journal of Metodoloˇski zvezki,: 1*(1) 205-212

Xu, R., Vaida, F & Harrington, D.P (2009). Using Profile Likelihood for Semiparametric Model Selection with Application to Proportional Hazards Mixed Models. *Statistica Sinica 19*, 819-842

Yamaguchi, K. (1991). *Event history analysis*. London: Sage Publications.

Yang, H., Chen, I.C., Tsay, Y.C., Zheng-Rong Li, Z.R., Chen,C., Hwu, H., & Chen-Hsin Chen, C.H. (2017). Using an Event-History with RiskFree Model to Study the Genetics of Alcoholism. *Scientific reports, 7,* 1975. DOI:10.1038/s41598-017-01791-4

Yi-lang, T., Xiang, X., Wang, X., Cubells, J.F., Babor, T.F., & Hao, W (2013). Alcohol and alcohol-related harm in China: policy changes needed. *Journal of Bull World Health Organization, 91*, 270–276. Retrieved from https://scholargoogle.com

Youngson, R.M (1999). *Collins Dictionary of Medicine*. London: Collins Pub.

Zucker, R.A., Boyd, G.M., and Howard, J., (Eds.). (1994). Development of Alcohol Problems: Exploring Biopsychosocial Matrix of risk. NIAAA Research Monograph No. 26. NIH Pub. No. 94-3495. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Zucker, R.A., Ellis, D.A., Bingham, C.R., & Fitzgerald, H.E. (1996).The development of alcoholic subtypes: Risk variation among alcoholic families during the childhood years. *Alcohol Health Res World, 20*(1), 46–54

APPENDICES

Appendix A: Disease Conditions Associated With Alcohol As Coded In ICD-10

Disease conditions which are by definition alcohol-attributable (AAF = 100%)

- ✓ E24.4 Alcohol-induced pseudo-Cushing's syndrome
- ✓ F10 Mental and behavioural disorders due to use of alcohol
- ✓ F10.0 Acute intoxication
- ✓ F10.1 Harmful use
- ✓ F10.2 Dependence syndrome
- ✓ F10.3 Withdrawal state
- ✓ F10.4 Withdrawal state with delirium
- ✓ F10.5 Psychotic disorder
- ✓ F10.6 Amnesic syndrome
- ✓ F10.7 Residual and late-onset psychotic disorder
- ✓ F10.8 Other mental and behavioural disorders
- ✓ F10.9 Unspecified mental and behavioural disorder
- ✓ G31.2 Degeneration of nervous system due to alcohol
- ✓ G62.1 Alcoholic polyneuropathy
- ✓ G72.1 Alcoholic myopathy
- ✓ I42.6 Alcoholic cardiomyopathy
- ✓ K29.2 Alcoholic gastritis
- ✓ K70 Alcoholic liver disease
- ✓ K70.0 Alcoholic fatty liver
- ✓ K70.1 Alcoholic hepatitis
- ✓ K70.2 Alcoholic fibrosis and sclerosis of liver
- ✓ K70.3 Alcoholic cirrhosis of liver
- ✓ K70.4 Alcoholic hepatic failure
- ✓ K70.9 Alcoholic liver disease, unspecified
- ✓ K85.2 Alcohol-induced acute pancreatitis
- ✓ K86.0 Alcohol-induced chronic pancreatitis
- ✓ O35.4 Maternal care for (suspected) damage to foetus from alcohol
- ✓ P04.3 Foetus and new-born affected by maternal use of alcohol
- ✓ Q86.0 Foetal alcohol syndrome (dysmorphic)

- ✓ R78.0 Finding of alcohol in blood
- ✓ T51 Toxic effect of alcohol
- ✓ T51.0 *Ethanol*
- ✓ T51.1 *Methanol*
- ✓ T51.8 *Other alcohols*
- ✓ T51.9 *Alcohol unspecified*
- ✓ X45 Accidental poisoning by and exposure to alcohol
- ✓ X65 Intentional self-poisoning by and exposure to alcohol
- ✓ Y15 Poisoning by and exposure to alcohol, undetermined intent
- ✓ Y90 Evidence of alcohol involvement determined by blood alcohol level

Note: ICD codes in italics represent sub-codes within a main code of classification.

Table A.1 Mortality and morbidity of alcoholics in Kenya 2013-2019

morbidity					mortality			
year	females	males	**total**	time	year	Males	Females	**Grand total**
2013				0	2013	1318	278	**1596**
2014	4	5	**9**	1	2014	1579	339	**1918**
2015	8	11	**19**	2	2015	1474	291	**1765**
2016	4	10	**14**	3	2016	1073	265	**1338**
2017	60	201	**261**	4	2017	1152	217	**1369**
2018	955	3136	**4091**	5	2018			
2019	495	1742	**2237**	6	2019			
					Total	**6596**	**1390**	**7986**

Table.A2: Reported diagnosed cases of alcoholism from 2014 to September 2019

Year	2014: Q1	2014: Q2	2014: Q3	2014: Q4	2015: Q1	2015: Q2	2015: Q3	2015: Q4
Total	1	1	4	3	2	1	8	8
Time	1	2	3	4	5	6	7	8

year	2016: Q1	2016: Q2	2016: Q3	2016: Q4	2017: Q1	2017: Q2	2017: Q3	2017: Q4
total	2	4	2	6	14	21	10	216
Time	9	10	11	12	13	14	15	16

year	2018:Q1	2018:Q2	2018:Q3	2018:Q4	2019:Q1	2019:Q2	2019:Q3
total	1005	1014	1070	1002	835	787	607
Time	17	18	19	20	21	22	23

Appendix B: Traditional Sir Model And Proposed Model Of Alcoholism

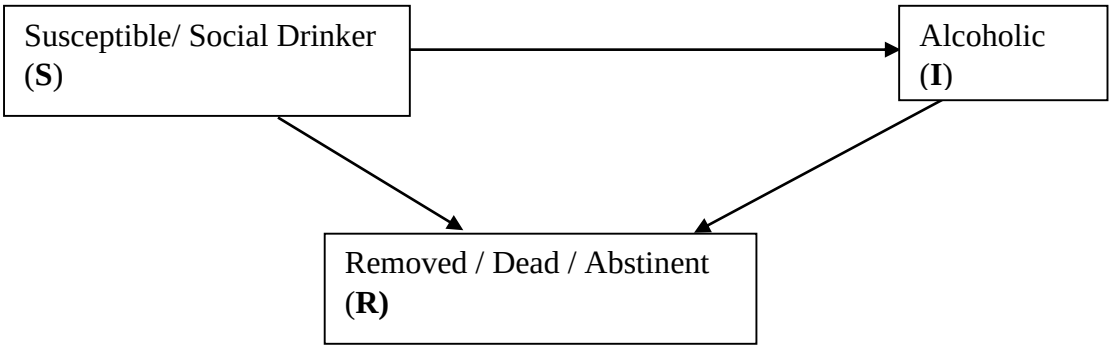


Figure AB.1 Traditional SIR alcoholism model showing the spread of alcoholism through interaction

Progression/ stages to alcoholism diagram.

This study focused on the process figuratively represented below

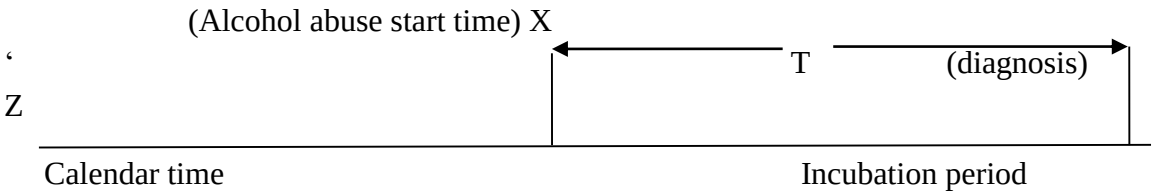


Figure AB.2 Progression to alcoholism showing the incubation period of alcoholism

Appendix C: Graphical display of effects of factors on hazard.

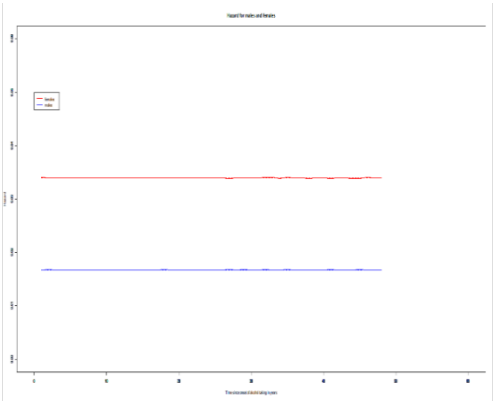


Fig 1. Effect of adding age onset

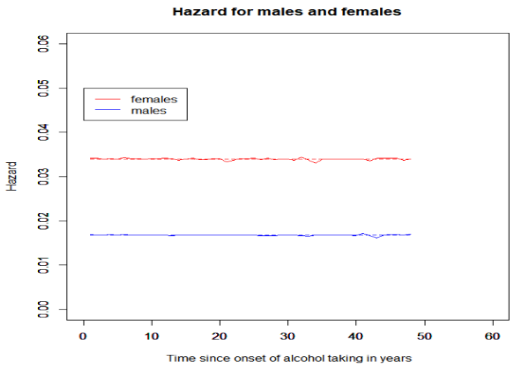


Fig 2. Effect of adding socio-cultural

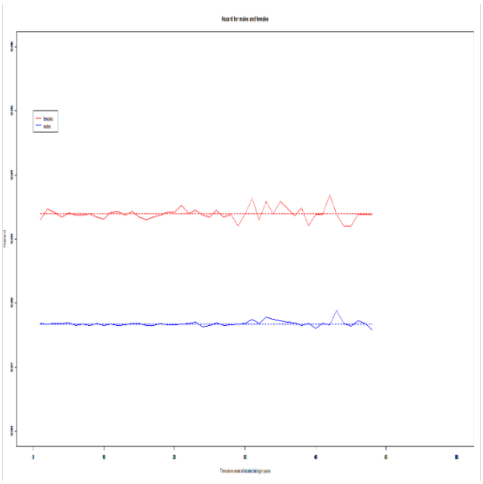


Fig 3. Effect of adding peer influence

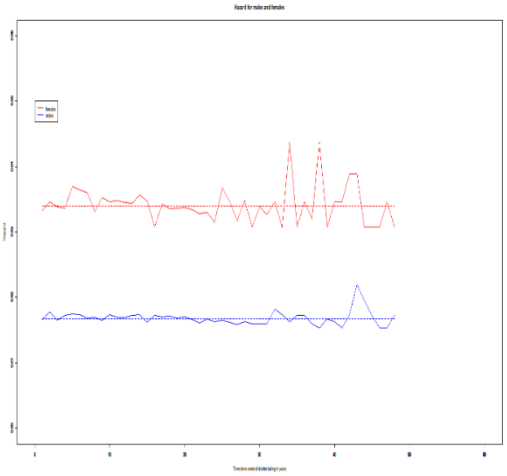


Fig 4. Effect of adding environment

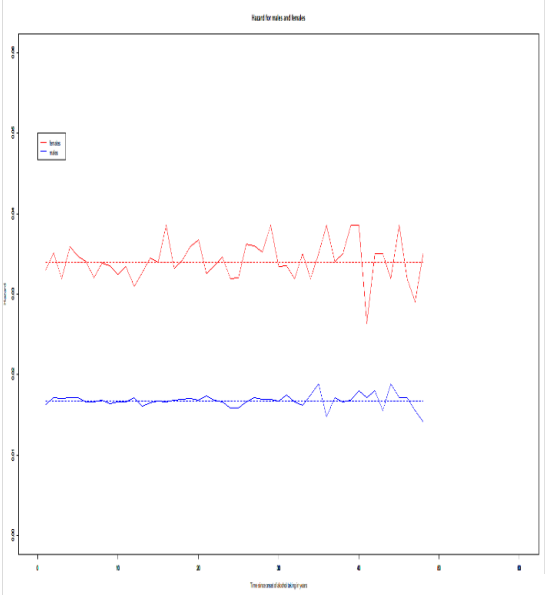


Fig 5. Effect of adding drinking habits and patterns

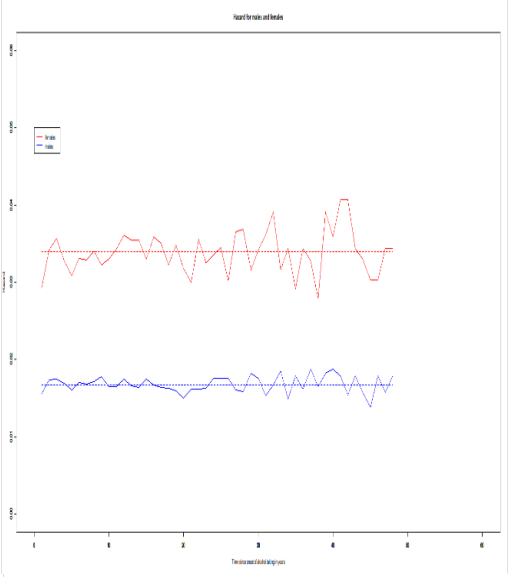
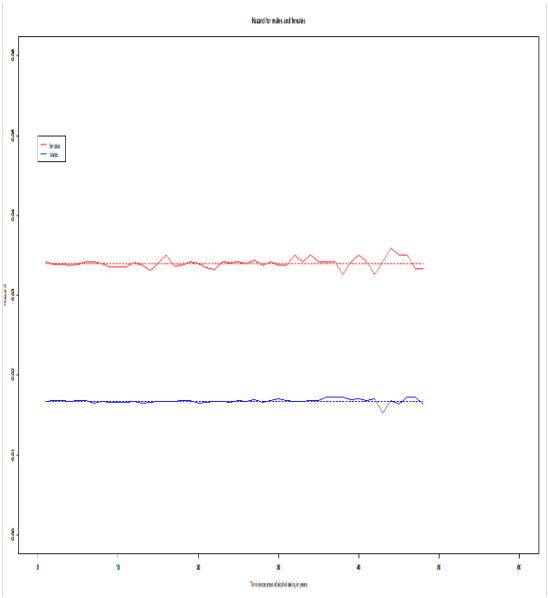


Fig 6. Effect of adding personality



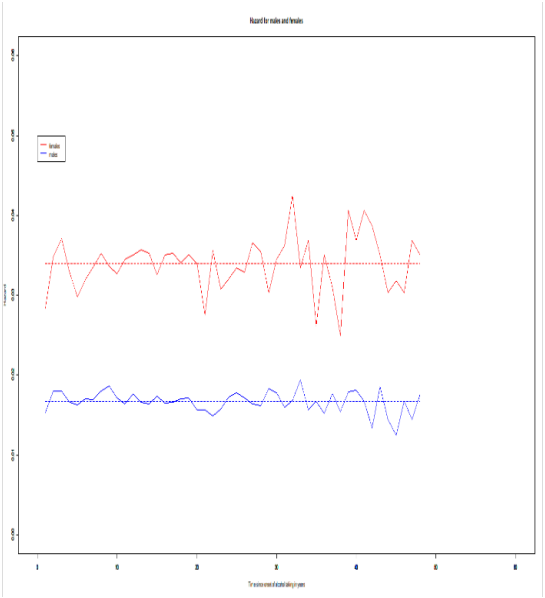


Fig 7. Effect of adding economic status Fig 8. Effect of adding family attention

Appendix D. Data Collection Matrix: Sampled Counties and Sub-Counties

- The conceptual SEM model had 12 observable variables. If p is the number of observed variable in a model, then the number of estimated parameters cannot be more than p(p+1)/2. Civelek (2018). 12(13)/2 =78. The sample size was 10 times the number of parameters that could be estimated, (Jayaram, Kannan, & Tan, 2004). 78x10=780 sample size
- The pilot sample size of 200 was based on the minimum size of a sample in SEM, (Celik & Yılmaz, 2013).

Rural areas

County	Sub-counties					Total
Nakuru	Gilgil	Molo	Subukia	Rongai	Naivasha	35
Nyeri	Mathira	Kieni	Othaya	Municipality	mukurweini	35
Kiambu	Thika	Limuru	Kikuyu	Githunguri	Kiambaa	35
Kirinyaga	Central	Mwea	Ndia	Gichugu		35
Nyandarua	Engineer	Ndaragwa	Kipipiri	Ol kalau	Kinangop	35
Uasin Gishu	Kesses	Soy	Moiben	Ainakboi	Turbo	35
						210

Urban centres

County	Main towns or sub-counties								Total
Nairobi	Kibra	Eastleigh	CBD	Kayole	Majengo	Bururburu	kawangware	Karen/langata	100
Nakuru	Nakuru	Molo	Naivasha	Gilgil	Subukia	Njoro	Maimahiu	Salgaa	70
Nyeri	Nyeri	Karatina	Kirichu	Mweiga	Othaya	Chaka	Mukurweini	kiamariga	70
Kiambu	Kiambu	Limuru	Kiambaa	Ruiru	Githurai	Juja	Githunguri	Kikuyu	70
Kirinyaga	Kerugoya	Sagana	Kutus	Kagio	Kagumo	Baricho	Kianyaga	Kianjagi	70
Nyandarua	Olkalau	Engineer	Njabini	Flyover	Shamata	Gwa kungu	Wiyumiririe	nyahururu	70
Uasin Gishu	Eldoret	Kesses	Burnt Forest	Moi's Bridge	Matunda	Turbo	Chepiteret	Jua kali	70
	1	2	3	4	5	6	7		520

Rehabs Total Respondents pecounty

Serial	County	Males	Females	Total		County	Number of respondents
1	Nakuru	8	2	10		Nakuru	115
2	Nyeri	4	1	5		Nyeri	110
3	Kiambu	8	2	10		Kiambu	115
4	Nyandarua	4	1	5		Nyandarua	110
5	Kirinyaga	4	1	5		Kirinyaga	110
6	Uasin Gishu	3	2	5		Uasin Gishu	110
7	Nairobi	8	2	10		Nairobi	110
Total		39	11	50		**Total**	**780**

Appendix E: Questionnaire on Alcoholism.

THE QUESTIONNAIRE. **Serial No……**

Age and Gender

1. Gender: ☐ M F ☐
2. Age: ……

   <15 ☐ 16-19 ☐ 20-24 ☐ 25-30 ☐ 31-35 ☐ 36-45 ☐ >45 ☐
3. Race: African ☐ Others ☐
4. Religion.

   Christian (Catholic) ☐ Christian (Protestant) ☐ Muslim ☐ Others ☐
5. Classify the neighbourhood you live in.

   low or slums ☐ middle ☐ Very rich ☐
6. State your highest education level

   No formal education ☐

   Completed primary school or less ☐

   Partially attended secondary school ☐

   Completed secondary school ☐

   Partially attended college or university ☐

   Completed college or degree ☐
7. What is your marital status?

   Single ☐ separated or divorced ☐ widowed ☐ Married ☐
8. When did you last take alcoholic drink?

   Within this week ☐ within last week ☐ within 30 days but more than 1 week ago ☐ within the last year but not in the 30 days ago ☐ Not in past year ☐
9. What do your parents/ guardians do for a living? Mark only one box.

   Domestic worker

   ☐

Farmer ☐

Employed or self-employed ☐

Own business ☐

10. At what age did you take alcoholic drink for the first time? :…….

<12 ☐ 13-15 ☐ 16-19 ☐ 20-24 ☐ 25-30 ☐ 31-35 ☐ >35 ☐

11. State the highest education level for your parent or guardian

Completed primary school or less ☐

Partially attended secondary school ☐

Completed secondary school ☐

Partially attended college or university ☐

Completed college or degree ☐

12. How often do you take alcoholic beverage?

Every day ☐ 3-4 times a week ☐ Once or twice a week ☐

Once a month ☐ Occasionally ☐

13. What type of alcoholic drink do you normally take?

Beer ☐ Spirits ☐ Second generation spirits ☐ wine ☐ others ☐ name …..

14. During the last 30 days, how many drinks did you normally take in a single occasion?

>12 ☐ 8-11 ☐ 5-7 ☐ 3-4 ☐ 1-2 ☐ None ☐

15. How many alcohol selling outlets are near (less than 1 km) your place of residence?

More than 10 ☐ 5-10 ☐ 1-4 ☐ None ☐

Family, friend, experience and responsibilities. Tick one

	Yes	No
16. Most of my friends drink when we are together.		
17. I approve of my peers drinking behaviour		
18. I drink as a way to celebrate, socialize or for fun		

19. I love the feeling of being drunk/ high		
20. My parent/guardian/siblings drink alcohol		
21. Family gatherings involve alcohol taking.		
22. I see alcohol adverts during watershed hour		
23. Community has occasions/ceremonies where alcohol taking by all is permitted		
24. I ignore warning about dangers of excessive drinking of alcohol		
25. Alcohol is available at home		
26. My best friend drink alcohol		
Give your views about some matters on why you continue taking alcohol	Yes	No
27. I don't fear getting in the 'risky' behaviour		
28. Excessive alcohol taking runs in our family		
29. I approve my friends drinking behaviour		
30. I have stress or difficulties in at work place		
31. Drinking alcohol promotes self-confidence		
32. Drinking excess alcohol shows you are a hero		
33. Alcohol related laws and enforcement are weak		
34. Job/occupation/daily activity expose me to alcohol		

Give your view on the following issues about you and where you spend most of the time.

	Yes	No
35. Your house has basic amenities such as clean piped water and electricity		
36. I am employed or in business		
37. There is happiness, peace, fun or joy at home		
38. I am involved in sports		

39. I am active in religious activities		
40. We live in a permanent house		
41. Do you have problems with alcohol?		
42. Has a doctor or family member advised you to stop drinking alcohol		

43. How old were you when you were diagnosed as having problems with alcohol? ……

Never ☐ <15 ☐ 16-19 ☐ 20-24 ☐ 25-30 ☐ 31-35 ☐ >35 ☐

The end. Thank you

Appendix E1: Key to coding used during analysis and model formation

S/N	Acronym	Item No	measurable variable represented in the questionnaire
1	act_rel		I am active in religious activities
2	adalc		I see alcohol adverts during watershed hour
3	alc_hme		Alcohol is available at home
4	ageonset		At what age did you take alcoholic drink for the first time?
5	alc_law		Alcohol related laws and enforcement are weak
6	am_hse		Your house has basic amenities such as clean piped water and el ectricity
7	app_frd		I approve my friends drinking behaviour
8	bsf_drn		My best friend drink alcohol
9	cel_soc		I drink as a way to celebrate, socialize or for fun
10	com_cer		Community has occasions/ceremonies where alcohol taking by all is permitted
11	dang_ex		I ignore warning about dangers of excessive drinking of alcohol
12	emp_bis		I am employed or in business
13	fam_gath		Family gatherings involve alcohol taking
14	fe_risk		I don't fear getting in the 'risky' behaviour
15	fred		Most of my friends drink when we are together.
16	freq		How often do you take alcoholic beverage?
19	Gender		Gender
20	gene		Excessive alcohol taking runs in our family
21	hap_hse		There is happiness, peace, fun or joy at home
22	hero_ex		Drinking excess alcohol shows you are a hero
23	high_edc		State your highest education level
24	hse_type		We live in a permanent house
25	job_ex		Job/occupation/daily activity expose me to alcohol
26	joints		How many alcohol selling outlets are near (less than 1 km) your p lace of residence?
27	las_drnk		When did you last take alcoholic drink?
28	lov_feel		I love the feeling of being drunk/ high
29	mar_sta		What is your marital status?
30	neighb		Classify the neighbourhood you live in
31	p_educ		State the highest education level for your parent or guardian
32	pa_liv		What do your parents/ guardians do for a living?
33	peer_dr		I approve of my peers drinking behaviour
34	quanty		During the last 30 days, how many drinks did you normally take in a single occasion?
35	race	3	Race
36	relign	4	Religion
37	self_con		Drinking alcohol promotes self-confidence
38	sib_drn		My parent/guardian/siblings drink alcohol
39	sport		I am involved in sports
40	stre_wrk		I have stress or difficulties in at work place
41	typealc		What type of alcoholic drink do you normally take?

Appendix F: Effects Of High Risk Drinking The Body Of A Drinker

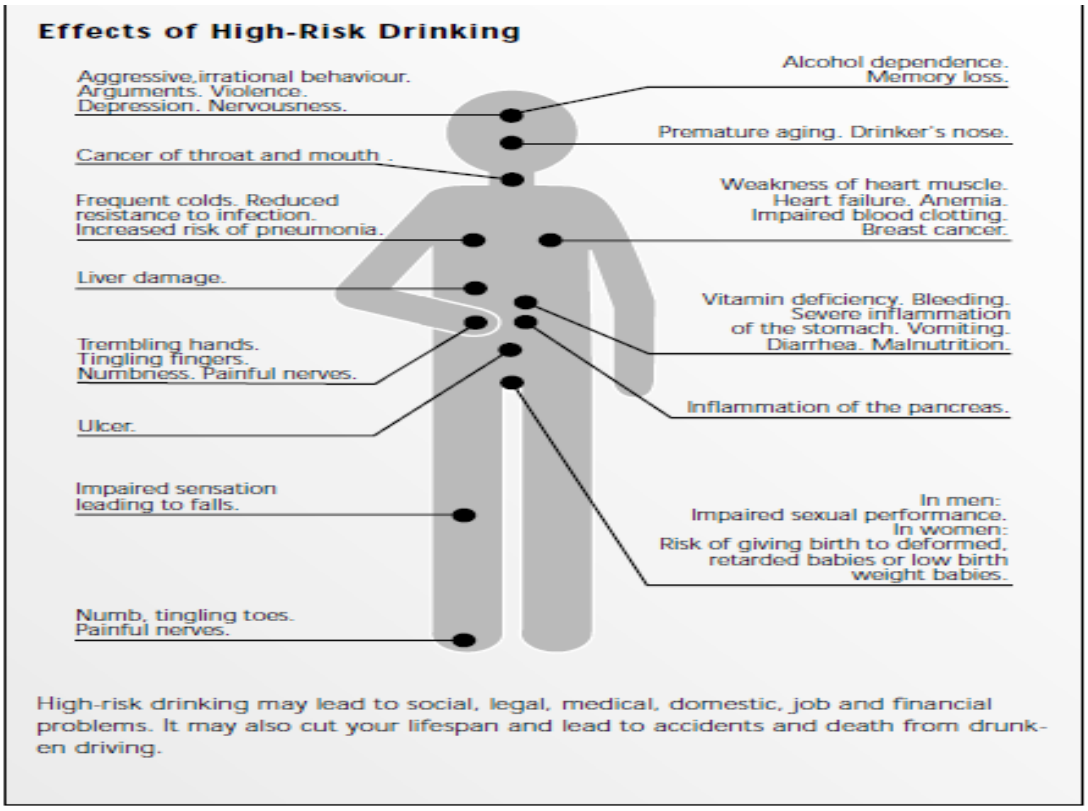


Figure AF.1 Effects of high risk drinking. (Adapted from Nacoa, n.d.)

Appendix G: Multilevel Models Of Alcoholism.

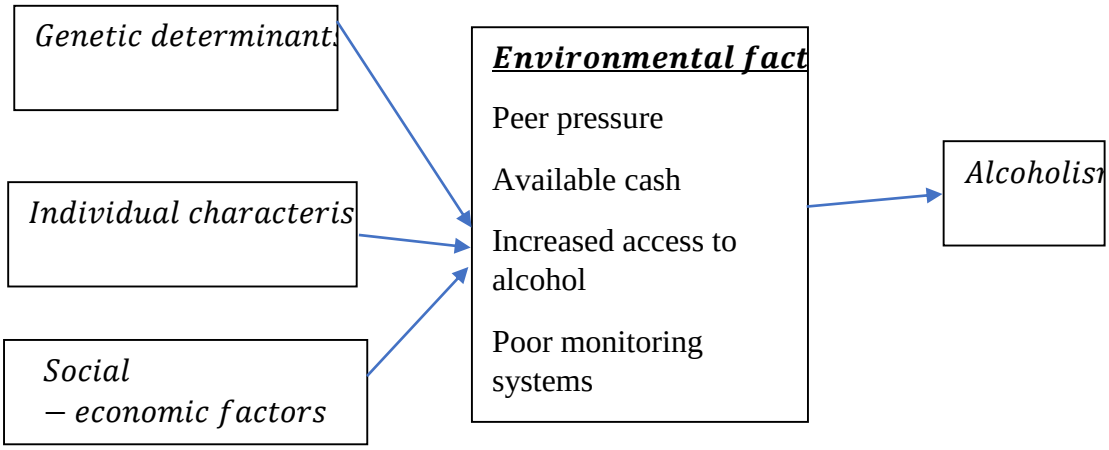
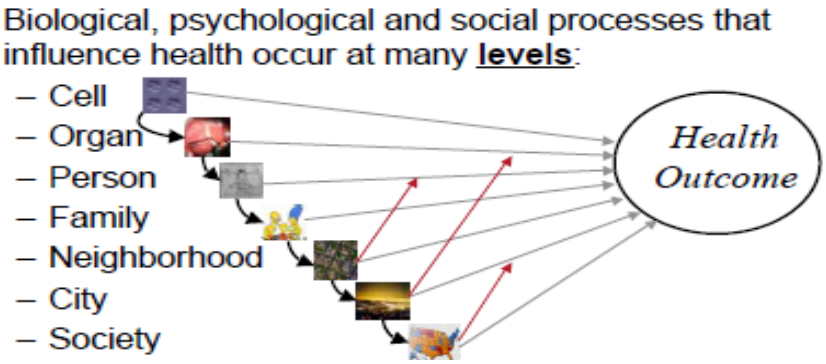


Figure G1. Hierarchical model. Multi-level model Risk factors of alcohol initiation with environment playing a modifier role. adapted from Hassan (2013)



Level:	
1. Cell:	Neurochemistry
2. Organ:	Ability to metabolize ethanol
3. Person:	Genetic susceptibility to addiction
4. Family:	Alcohol abuse in the home
5. Neighborhood:	Availability of bars
6. Society:	Regulations; organizations; social norms

Figure G2. Six level model, adapted from “Lecture 1” (n.d.).

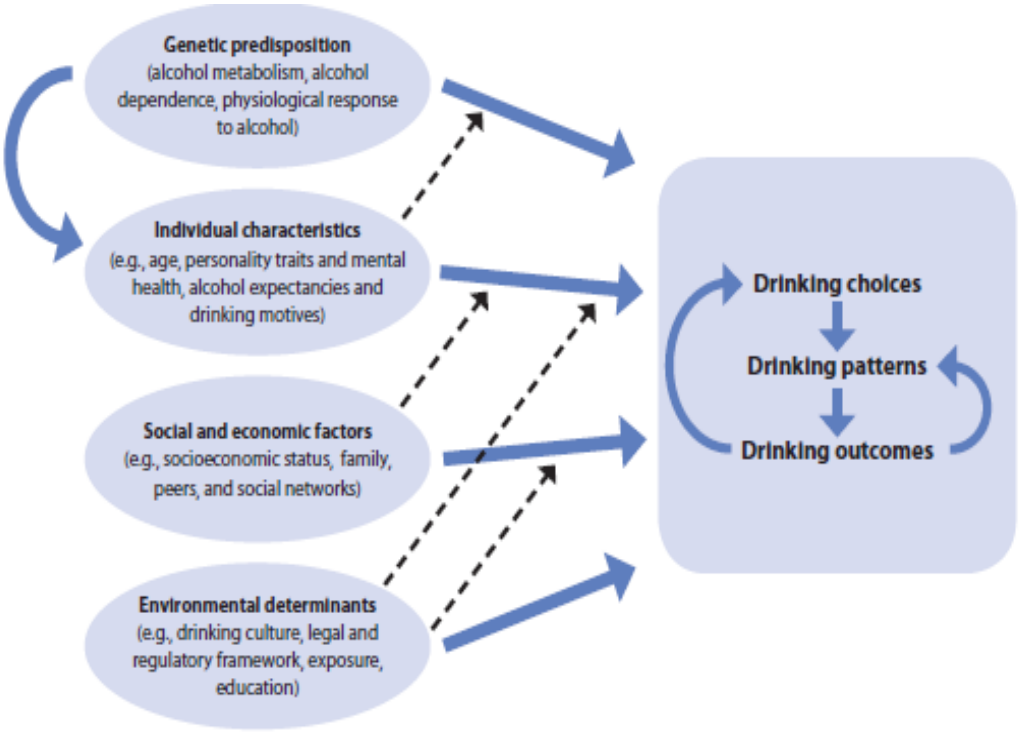


Figure G3: Model of interactions between factors adapted from ICAP (2009)

Appendix H. Typical results of model fitting.

Factor	Estimate	Std. Error	z value	Pr(>|z|)	
factor(PERIOD)0	0.3365	0.5855	0.575	0.56554	
factor(PERIOD)1	-1.0986	0.3651	-3.009	0.00262	**
factor(PERIOD)2	-1.8281	0.3591	-5.091	3.57e-07	***
factor(PERIOD)3	-2.4756	0.3471	-7.133	9.83e-13	***
factor(PERIOD)4	-2.3026	0.2803	-8.215	< 2e-16	***
factor(PERIOD)5	-3.0540	0.3236	-9.438	< 2e-16	***
factor(PERIOD)6	-3.0301	0.3238	-9.359	< 2e-16	***
factor(PERIOD)7	-2.9565	0.2845	-10.393	< 2e-16	***
factor(PERIOD)8	-3.1931	0.3076	-10.379	< 2e-16	***
factor(PERIOD)9	-3.3282	0.3393	-9.810	< 2e-16	***
factor(PERIOD)10	-3.2611	0.2826	-11.539	< 2e-16	***
factor(PERIOD)11	-3.4063	0.2819	-12.083	< 2e-16	***
factor(PERIOD)12	-3.5020	0.3060	-11.444	< 2e-16	***
factor(PERIOD)13	-3.8686	0.3572	-10.830	< 2e-16	***
factor(PERIOD)14	-3.7471	0.3199	-11.712	< 2e-16	***
factor(PERIOD)15	-4.0518	0.4118	-9.840	< 2e-16	***
factor(PERIOD)16	-3.8395	0.3196	-12.013	< 2e-16	***
factor(PERIOD)17	-4.0452	0.3566	-11.343	< 2e-16	***
factor(PERIOD)18	-4.6559	0.4493	-10.362	< 2e-16	***
factor(PERIOD)19	-4.6603	0.5801	-8.034	9.44e-16	***
factor(PERIOD)20	-4.9813	0.5793	-8.598	< 2e-16	***
factor(PERIOD)21	-4.0839	0.3565	-11.455	< 2e-16	***
factor(PERIOD)22	-4.5721	0.5026	-9.097	< 2e-16	***
factor(PERIOD)23	-4.9628	0.5794	-8.566	< 2e-16	***
factor(PERIOD)24	-4.8853	0.5795	-8.430	< 2e-16	***
factor(PERIOD)25	-4.4705	0.4498	-9.940	< 2e-16	***
factor(PERIOD)26	-4.5850	0.5803	-7.901	2.76e-15	***
factor(PERIOD)27	-4.0073	0.5045	-7.943	1.98e-15	***
factor(PERIOD)28	-4.6102	0.7106	-6.488	8.72e-11	***
factor(PERIOD)29	-5.0999	0.7093	-7.190	6.46e-13	***
factor(PERIOD)30	-5.5094	1.0020	-5.498	3.84e-08	***
factor(PERIOD)31	-4.1431	0.7127	-5.813	6.12e-09	***
factor(PERIOD)32	-4.8752	1.0038	-4.857	1.19e-06	***
factor(PERIOD)33	-4.3610	0.5810	-7.506	6.11e-14	***
factor(PERIOD)34	-4.2341	1.0072	-4.204	2.63e-05	***
factor(PERIOD)35	-4.6728	1.0047	-4.651	3.30e-06	***
factor(PERIOD)36	-5.3982	1.0023	-5.386	7.20e-08	***
factor(PERIOD)37	-3.9053	0.5831	-6.697	2.13e-11	***
factor(PERIOD)38	-4.1589	0.5818	-7.148	8.82e-13	***
factor(PERIOD)39	-3.3673	0.5085	-6.621	3.56e-11	***
factor(PERIOD)40	-2.7344	0.4615	-5.925	3.13e-09	***
factor(PERIOD)41	-3.1864	0.4564	-6.982	2.91e-12	***
factor(PERIOD)42	-3.7377	0.7155	-5.224	1.75e-07	***
factor(PERIOD)43	-4.4659	1.0057	-4.440	8.98e-06	***

factor(PERIOD)44	-2.6391	0.5976	-4.416	1.01e-05 ***
factor(PERIOD)45	-16.5661	353.7936	-0.047	0.96265
factor(PERIOD)46	-4.5326	1.0054	-4.508	6.53e-06 ***
factor(PERIOD)47	-3.8501	0.7146	-5.388	7.13e-08 ***
factor(PERIOD)53	-3.9703	1.0094	-3.933	8.38e-05 ***
factor(PERIOD)54	-3.9890	1.0092	-3.953	7.73e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 16295.9 on 11755 degrees of freedom
Residual deviance: 2334.3 on 11705 degrees of freedom
AIC: 2434.3

Number of Fisher Scoring iterations: 15

Appendix J: Results of fitting glm model for age at year of diagnosis

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.8930	-0.0152	-0.0152	-0.0152	4.2578

Coefficients:				
	Estimate	Std. Error	z value	Pr(>|z|)
years.diagnosis10	-17.5661	2797.4419	-0.006	0.994990
years.diagnosis11	-17.5661	1318.7268	-0.013	0.989372
years.diagnosis12	-17.5661	688.6826	-0.026	0.979651
years.diagnosis13	-17.5661	548.6235	-0.032	0.974457
years.diagnosis14	-3.1781	0.5893	-5.393	6.92e-08 ***
years.diagnosis15	-3.9703	0.7137	-5.563	2.66e-08 ***
years.diagnosis16	-4.7185	1.0045	-4.698	2.63e-06 ***
years.diagnosis17	-3.2809	0.4555	-7.202	5.92e-13 ***
years.diagnosis18	-3.9640	0.5828	-6.802	1.03e-11 ***
years.diagnosis19	-2.1001	0.2497	-8.410	< 2e-16 ***
years.diagnosis20	-2.5366	0.3132	-8.099	5.56e-16 ***
years.diagnosis21	-2.9178	0.3630	-8.038	9.11e-16 ***
years.diagnosis22	-3.1282	0.4171	-7.500	6.38e-14 ***
years.diagnosis23	-2.5719	0.3128	-8.222	< 2e-16 ***
years.diagnosis24	-2.3671	0.3019	-7.841	4.48e-15 ***
years.diagnosis25	-2.0065	0.2663	-7.536	4.86e-14 ***
years.diagnosis26	-2.9293	0.3879	-7.551	4.32e-14 ***
years.diagnosis27	-3.1527	0.4567	-6.904	5.06e-12 ***
years.diagnosis28	-2.0794	0.2942	-7.069	1.56e-12 ***
years.diagnosis29	-1.8718	0.2871	-6.520	7.03e-11 ***
years.diagnosis30	-1.2809	0.2528	-5.068	4.03e-07 ***
years.diagnosis31	-1.6422	0.3154	-5.207	1.92e-07 ***
years.diagnosis32	-1.9459	0.3780	-5.148	2.63e-07 ***
years.diagnosis33	-1.3218	0.3249	-4.068	4.74e-05 ***
years.diagnosis34	-2.3273	0.5238	-4.443	8.88e-06 ***

years.diagnosis35	-1.2528	0.3586	-3.494	0.000476 ***
years.diagnosis36	-2.6027	0.5182	-5.023	5.10e-07 ***
years.diagnosis37	-1.6864	0.4869	-3.464	0.000533 ***
years.diagnosis38	-1.7492	0.5417	-3.229	0.001243 **
years.diagnosis39	-1.5581	0.5501	-2.832	0.004620 **
years.diagnosis40	-0.3185	0.4647	-0.685	0.493125
years.diagnosis41	-0.5596	0.6268	-0.893	0.371944
years.diagnosis42	-1.7918	1.0801	-1.659	0.097147 .
years.diagnosis43	-17.5661	1615.104	-0.011	0.991322
years.diagnosis44	1.6094	1.0954	1.469	0.141776
years.diagnosis45	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis46	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis47	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis48	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis49	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis50	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis51	-17.5661	3956.1803	-0.004	0.996457
years.diagnosis52	17.5661	3956.1803	0.004	0.996457
years.diagnosisC	-9.0644	1.0001	-9.064	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
Null deviance: 15850.9 on 11434 degrees of freedom
Residual deviance: 1471.3 on 11390 degrees of freedom
AIC: 1559.3

Appendix K: First publication emanating from this study

Int. J. Agricult. Stat. Sci. Vol. 17, No. 2, pp. 545-556, 2021
DocID: https://connectjournals.com/03899.2021.17.545
www.connectjournals.com/ijass
ISSN : 0973-1903, e-ISSN : 0976-3392
ORIGINAL ARTICLE

MODELLING HAZARD OF BECOMING ALCOHOLIC USING PARAMETRIC AND NON-PARAMETRIC METHODS

George Mwangi Muriuki*, John M. Mutiso and Mathew K. Kosgei

Department of Mathematics, Physics and Computing, Moi University, 3900-30100, Eldoret, Kenya
E-mail: muriuki.gmm@gmail.com

Abstract: Hazard function shows the dependence of chance of onset of alcoholism on time. A cross-sectional study applying multistage stratified sampling method was conducted using questionnaire administered in seven out of 47 counties to current alcohol consumers and persons in rehabilitation centres which captured latent variables: peer influence, family attention, personality, economic status, environmental setting, socio-cultural, drinking habits and patterns. Parametric hazard function was determined using Birnbaum-Saunders (B-S) model. Person-period dataset and discrete-time proportional hazards model were used to determine non-parametric hazard rates. Parametric hazard function based was an upside-down curve, while failure rate was an increasing curve. Non-parametric hazard of alcoholism is an upside-down curve with bumps and valleys. The non-parametric failure rate of alcoholism based on all factors was an increasing upward curve. Probability of becoming alcoholic increased from 0.31% when drinking once per week to 57% when drinking seven times a week. Based on years since onset of drinking it rose from 0.0059% in year one to 70% after 30 years. Thus, probability of becoming alcoholic increased with frequency of drinking per week and with time since onset of alcohol taking. Females had higher hazard of becoming alcoholic than males and was affected by factors differently. B-S distribution was confirmed suitable to model time from onset to diagnosis by upside-down parametric hazard function. Hazard of alcoholism is highly affected by gender, personality and peer influence. Anti-alcoholism programs to focus in reducing the frequency of drinking.

Key words: Hazard rate, Alcoholism, Risk factor, Event-history analysis, Birnbaum-Saunders distribution.

Cite this article

George Mwangi Muriuki, John M. Mutiso and Mathew K. Kosgei (2021). Modelling Hazard of becoming Alcoholic using Parametric and Non-Parametric methods. *International Journal of Agricultural and Statistical Sciences*. DocID: https://connectjournals.com/03899.2021.17.545

Appendix L: Second publication based on this research

George M. Muriuki, 2021, 9:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

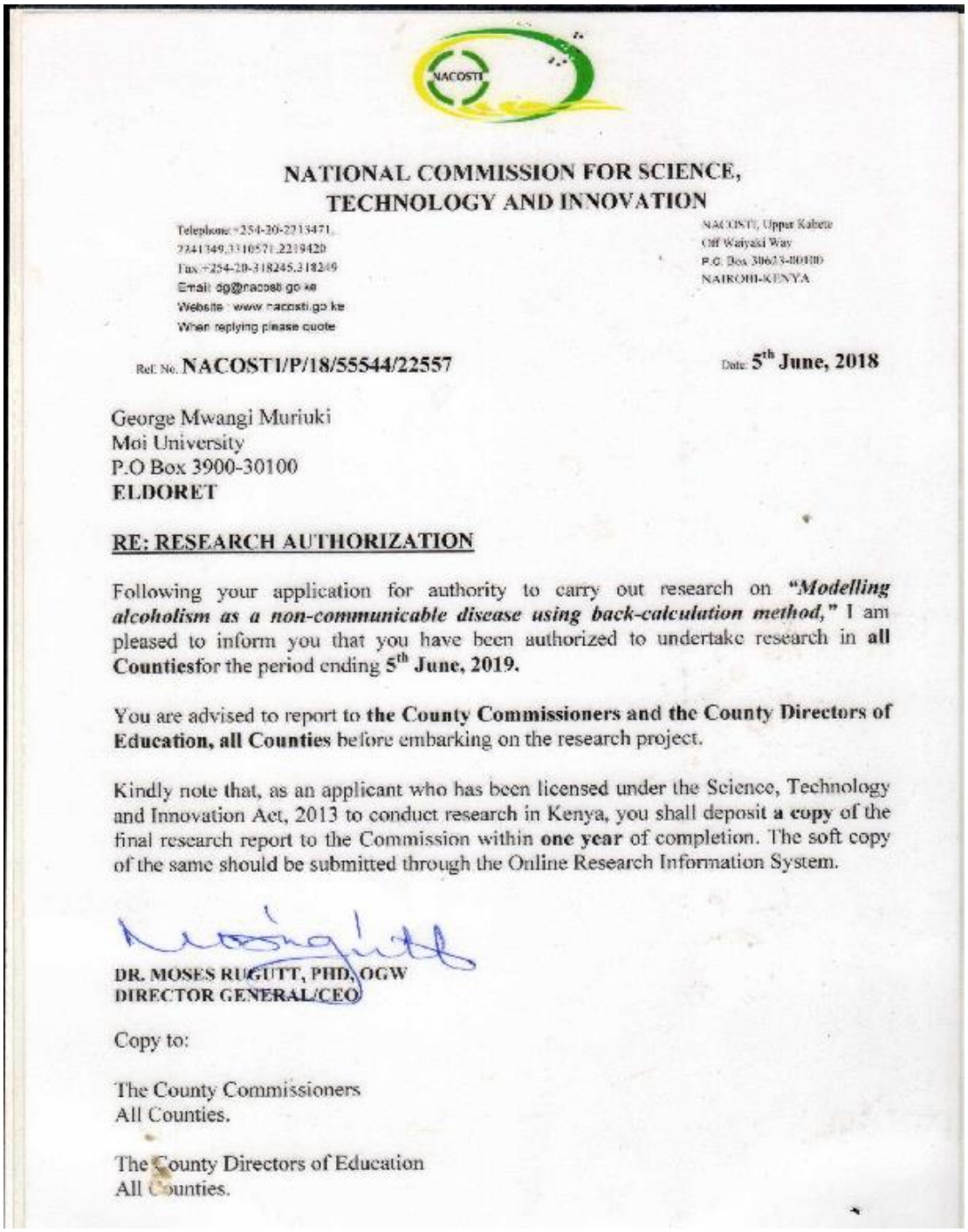
Modelling Risk factors of Alcoholism Using Structural Equation Model

Phd Scholar George M. Muriuki, Dr. Mathew K. Kosgei, Prof. John M. Mutiso
Department of Mathematics,
Physics and Computing, Moi University,
P.O Box, 3900-30100, Eldoret, Kenya.
muriuki.gmm@gmail.com, mkosgei12@gmail.com, johnkasome@yahoo.com

Abstract- Alcohol has become alarmingly the most routinely used and misused substance among both adults and minors in the contemporary society and its socio-economic impact is devastating. The risk factors strongly linked to alcoholism and many other problems related to alcohol use and misuse have been studied disjointedly. Different structural models illustrating relationship between alcoholism and associated risk factors have been proposed. The objective was to test if the structure of a proposed theoretical hierarchical model comprised of specific observed variables clustered to measure nine latent variables. Structural equation modelling (SEM) was used to test the direct effect of the latent variables of personality, gender, age at onset of alcohol taking, peer influence, drinking habits and patterns, family attention, economic status, environment and social-cultural on alcoholism. Subjects who comprised of inmates at the rehabilitation centers and current alcohol users were evaluated and completed a battery of psychosocial-related questionnaires. The initially proposed model structure was modified due to misspecifications. The modified model exhibited adequate global fit indexes (RMSEA = .06 (.056, .091), CFI =.803, SRMR=.06, $\chi 2$ (df) =1733(464) and acceptable measures of reliability. The nine factor (Gender, age at onset, drinking habits and patterns, family and family attention, economic status, personality, peer pressure, socio-cultural, and environment) alcoholism model fitted well. The findings suggest those parents' education and living conditions, drinker's education level, race, quantity drunk type of alcohol drunk and stress at work were significant predictors of alcoholism. Knowledge of the linkage between factors is useful in designing alcohol abuse preventive programs.

Keywords: Structural Equation Modelling, hierarchical models, latent variables, factors of alcoholism.

Appendix M: Permit to conduct this research



NATIONAL COMMISSION FOR SCIENCE,
TECHNOLOGY AND INNOVATION

Telephone: +254-20-2213471,
2241349,3310571,2219420
Fax: +254-20-318245,318249
Email: dg@nacosti.go.ke
Website : www.nacosti.go.ke
When replying please quote

NACOSTI, Upper Kabete
Off Waiyaki Way
P.O. Box 30623-00100
NAIROBI-KENYA

Ref. No. **NACOSTI/P/18/55544/22557**

Date: **5th June, 2018**

George Mwangi Muriuki
Moi University
P.O Box 3900-30100
ELDORET

RE: RESEARCH AUTHORIZATION

Following your application for authority to carry out research on ***"Modelling alcoholism as a non-communicable disease using back-calculation method,"*** I am pleased to inform you that you have been authorized to undertake research in **all Counties**for the period ending **5th June, 2019.**

You are advised to report to **the County Commissioners and the County Directors of Education, all Counties** before embarking on the research project.

Kindly note that, as an applicant who has been licensed under the Science, Technology and Innovation Act, 2013 to conduct research in Kenya, you shall deposit **a copy** of the final research report to the Commission within **one year** of completion. The soft copy of the same should be submitted through the Online Research Information System.

DR. MOSES RUGUTT, PHD, OGW
DIRECTOR GENERAL/CEO

Copy to:

The County Commissioners
All Counties.

The County Directors of Education
All Counties.