# BAYESIAN APPROACH FOR JOINT MODELING OF TIME TO SEIZURE FREEDOM AND SEIZURE FREQUENCY COUNT IN EPILEPTIC PATIENT AT HAWASSA UNIVERSITY COMPREHENSIVE SPECIALIZED HOSPITAL

M.SC. THESIS

BY

YITAGESU ESHETU

HAWASSA UNIVERSITY, HAWASSA, ETHIOPIA

November, 2023

# BAYESIAN APPROACH FOR JOINT MODELING OF TIME TO SEIZURE FREEDOM AND SEIZURE FREQUENCY COUNT IN EPILEPTIC PATIENT AT HAWASSA UNIVERSITY COMPREHENSIVE SPECIALIZED HOSPITAL

MSC. Thesis

By: YITAGESU ESHETU

Advisor: Selamawit Serka Moja (PhD.)

A THESIS SUBMITTED TO COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE DEPARTMENT OF STATISTICS HAWASSA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTERS OF SCIENCES IN STATISTICS

November, 2023

HAWASSA, ETHIOPIA

# APPROVAL SHEET -1

This is to certify that the thesis entitled “Bayesian Approach for Joint Modeling of Time to Seizure Freedom and Seizure Frequency Count in Epileptic Patient at Hawassa University Comprehensive Specialized Hospital” submitted in partial fulfillment of the requirements for the degree of master of science in statistics with specialization in Applied Statistics of the graduate program of the College Of Natural and Computational Science, Hawassa University, and is a record of original research carried out by Yitagesu Eshetu Ayele ID.No: GpApStR/0007/14, under my supervision, and no part of the thesis has been submitted for any other degree or diploma. The guidance and assistance received during the course this investigation have all been duel acknowledged. Therefore, I recommend that it would be accepted fulfilling the thesis requirements.

| _____ | _____ | _____ |
|---|---|---|
| Name of Advisor | Signature | Date |

# APPROVAL SHEET -2

We the undersigned, member of the board of examiners of the final open defense by Yitagesu Eshetu have read and evaluated his thesis entitled “Bayesian Approach for Joint Modeling of Time to Seizure Freedom and Seizure Frequency Count in Epileptic Patient at Hawassa University Comprehensive Specialized Hospital” and examined the candidate. This is, therefore, to certify that thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Statistics (Specialization: Applied Statistics).

| | | |
|---|---|---|
| _____ | _____ | _____ |
| Name of Department Head | Signature | Date |
| _____ | _____ | _____ |
| Name of Advisor | Signature | Date |
| _____ | _____ | _____ |
| Name of External Examiner | Signature | Date |
| _____ | _____ | _____ |
| Name of Internal Examiner I | Signature | Date |
| _____ | _____ | _____ |
| Name of Internal Examiner II | Signature | Date |
| _____ | _____ | _____ |
| School of Graduate committee | Signature | Date |

# Table of Contents

# ACKNOWLEDGMENT

First and foremost, I am grateful to the Almighty God for His mercy and grace, which have enabled me to successfully complete my studies.

I would like to express my deepest gratitude to my advisor Selamawit Serka Moja (PhD), for her excellent guidance, instructions and constructive comments from the start of proposal development to the final work of this thesis. This thesis would not be possible without her advice, suggestions and encouragement.

I would like to extend my sincere thanks to Dembi Dollo University for providing me financial support without which this thesis would not have been possible.

I am deeply grateful to all staff members of Neurologic Clinic at Hawassa Comprehensive Specialized Referral Hospital and epileptic patients for their unreserved support and collecting the data precisely for this study.

Finally, I would like to thank Hawassa University staffs, friends and my loving and supportive parents, who directly or indirectly involved in this thesis.

**LIST OF ABBREVIATION**

AEDS ---------- Anti-Epileptic Drugs

AIC ----------Akaike Information Criteria

AMSH ----------Amanuel Mental Specialized Hospital

AOR ----------Adjusted Odds Ratio

ART---------- Antiretroviral Therapy

BIC ---------- Bayesian Information Criteria

EEG---------- Electroencephalogram

GLMM ---------- Generalized Linear Mixed Model

HIV/AIDS---------- Acquired Immunodeficiency Syndrome Human Immunodeficiency Virus

HUCSH ---------- Hawassa University Comprehensive Specialized Hospital

ILAE----------International League Against Epilepsy

MCMC ---------- Markov Chain Monte Carlo

MMAS ---------- Morisky Medication Adherence Scale

QoL---------- Quality of Life

WHO ---------- World Health Organization

# List of Tables

# List of Figures

# ABSTRACT

***Background**: Epilepsy was classified as a chronic, non-communicable brain condition by the World Health Organization. Over 50 million people globally have been impacted by epilepsy, one of the most common neurological disorders. In Ethiopia, epilepsy is one of the top 20 killers, and 5.2 out of every 1000 people will suffer from it at some point in their lifetime. The main objective of the study was to investigate predictors of seizure attacks progression and time to seizure freedom among epileptic patients using separate and joint analysis in Bayesian approach.*

***Methodology:** The study analyzed data from 203 epileptic patients who initiated anti-epileptic drugs (AEDs) at Hawassa University Comprehensive Specialized Hospital Neurologic Clinic between 1$^{st}$ May 2018 up to 1$^{st}$ May 2023. A retrospective cohort study design was carried out and epileptic patients age greater than 18 years old were used as source of population for this study and also the data obtained from HUCSH. A Bayesian approach for joint modeling is used to analysis time to seizure freedom and seizure frequency count.*

***Results**: Out of these patients, 80.3% (163) achieved seizure freedom, while 19.7% (40) were censored due to not achieving seizure freedom within the study period. Analyzing factors influencing seizure outcomes, the study found that Phenytoin usage showed a statistically significant positive effect on seizure reduction, while Phenobarbitone and Sodium Valproate did not exhibit significant effects. Having more treatment sessions had a significant positive effect on reducing seizures. Patients with a partial seizure type showed a significant increase in seizure frequency, while those who exercised, had a family history of epilepsy, or consumed alcohol experienced a significant reduction in seizure frequency. Patients without chronic diseases had significantly fewer seizures. Moreover, patients with co-morbidities or a history of alcohol consumption had a higher frequency of seizures.*

***Conclusion**: Bayesian joint modeling revealed that the Weibull survival model and Negative Binomial Zero-Inflated model provided the best fit for survival and count data, respectively. This study's findings contribute to a comprehensive understanding of the factors influencing seizure freedom and seizure frequency in epileptic patients, offering valuable insights for clinical management and treatment strategies.*

# 1. INTRODUCTION

## 1.1 Background of the Study

Epilepsy is classified as a chronic, non-communicable brain condition (WHO, 2022). Although it may afflict anybody at any age, epilepsy is more common in young children and older age groups. Men, especially teenagers, and the elderly were substantially more susceptible to seizures (Thijs et al., 2019 and Yin et al., 2021). The International League Against Epilepsy (ILAE) defined epilepsy as a brain disorder in 2014 if any of the following symptoms were present: One uncontrolled seizure and a diagnosis of an epilepsy syndrome; And at least two uncontrolled seizures occurring more than 24 hours apart; (Fisher et al., 2014).

Over 50 million people globally have been impacted by epilepsy, one of the most common neurological disorders. Those with epilepsy have a three times higher chance of dying before their natural death than the general population would (WHO and Thijs et al., 2019). Epilepsy prevalence across all ages rose from 17.0 million in 1990 to 27.3 million in 2017, a 60.6% rise. Age-standardized prevalence rates of epilepsy increased by 13.6% from 316.0 per 100,000 people in 1990 to 359.1 per 100,000 people in 2017, after adjusting for population and age structure (Yin et al., 2021). According to the WHO, up to 70% of epilepsy sufferers may avoid seizures if their condition was correctly identified and treated. People with epilepsy and their families experience stigma and prejudice in many areas of the world.

The epilepsy burden varied widely geographically among the 195 nations and territories included in the GBD 2017 survey. Several Asian and African nations with vast populations and poor socioeconomic position bear the brunt of the epilepsy burden. Lower socioeconomic level nations and impoverished areas typically have higher rates of epilepsy and wider gender gaps (Yin et al., 2021). According to WHO almost 80% of epilepsy sufferers reside in low- and middle-income nations. Even though, only 25% of epileptics in low-income nations receive the necessary medical care. Low-middle income nations and low-income countries, or 4.01% and 11.87% respectively, had greater prevalence ratios of uncontrolled epilepsy to treated epilepsy (Gagandeep et al., 2020).

In Ethiopia, epilepsy is one of the top 20 killers, and 5.2 out of every 1000 people will suffer from it at some point in their lifetime (Yazie et al., 2021 and Melkamu et al., 2021). Ethiopia,

where persons with epilepsy and their family experience a significant level of perceived stigma, has one of the highest rates of epilepsy among school-age children and teens (Meron et al., 2016). Despite the fact that in Ethiopia epilepsy is also closely linked to illiteracy and other indicators of poverty (Vaid et al., 2012 and Hassen et al., 2020).

Meta-analysis and Systematic Review show all age groups of patients receiving antiepileptic drugs in Ethiopia were reported to have an overall pooled prevalence of seizure freedom of 46%. Among reports of the major research in Ethiopia, there were significant variations in the prevalence of seizure freedom. The smallest prevalence of seizure freedom was 8% as reported by the study conducted at Ayder Comprehensive Specialized Hospital and Mekele Hospital (K. Gebre and A. Haylay 2018). The highest prevalence of seizure freedom (82.4%) was reported by the study conducted at Gondar University Teaching Hospital (E. M. Birru et al., 2016).

According to research by (Marion et al., 2020), 38% of patients experienced an poor uncontrolled seizure. Uncontrolled seizures were linked to low adherence to anti-seizure drugs (ASDs) [AOR: 2.04], having not much formal education [AOR: 2.71], and using ASDs for just a year [AOR: 4.90]. Uncontrolled seizures were seen in a notably larger percentage of epileptic individuals. Uncontrolled seizures were more common among epileptic individuals with minimal formal education, those who took their medicine for a brief period of time but failed to adhere to it, and those with poor medication adherence.

While the ultimate objective of treating epileptic patients is to keep them seizure-free, some people may still have seizures after receiving anti-epileptic medication. Thus, it is critical to identify seizure-provoking circumstances in epileptic patients receiving AED medication in order to enhance the patient's quality of life and enable them to do their actions as planned (Kebede et.al., 2022). This thesis aims to investigate factors affecting seizure attacks and seizure freedom of epileptic patients using separate and joint model application in the case of Hawassa University Comprehensive Specialized Hospital, Hawassa, Ethiopia.

### 1.2 Statement of the Problem

Epilepsy is a chronic brain disorder that affects people of all ages and is prevalent around the world (GBD 2016 Epilepsy Collaborators). Epilepsy has significant economic effects, especially in Africa, where it has placed a heavy load on already undeveloped healthcare systems in

developing countries like Ethiopia (Allers et al., 2015). Anti-elliptic medication non-adherence (AEMNA) prevalence was reported to be between 21.8 and 68% in Ethiopia. AEMNAs have a high burden due to a number of reasons, including a poor healthcare system, poor medical services, lack of drug access, financial limitations, negative impacts of antiepileptic drugs, and poor seizure controls (Belayneh et al., 2020).

In a recent study in Ethiopia, which involved 394 patients and was carried out at the Amanuel Mental Specialized Hospital (AMSH), approximately 15.74% of participants had a history of head injury caused by epilepsy, and 16.75% of them had a psychotic disorder. The study also revealed that factors such as age, smoking, chat use, lack of sleep, head injuries, depression, clumsiness, and treatment options all affect the number of seizure attack (Temam et al., 2021).

Keunbaik et al (2019) use marginalized models to simulate data on epileptic seizures that were gathered over time through repeated measurements. They found that there were more epileptic seizures in the AED arm compared with the placebo arm. Also, it was shown that the anticipated number of epileptic seizures decreased as Week increased. Similarly, Seizure episodes were recorded longitudinally using a zero-inflated modeling approach by Fenta et al (2019). Research showed that patients' lower socioeconomic status was linked to the frequency of seizure events. Low socioeconomic status, as measured by a lack of house ownership or a poor level of education, was also revealed to be a risk factor for adult-onset epilepsy.

Several studies use joint model of longitudinal and survival analysis to model number seizure and time to events. For instance, J. K. Rogers & J. L. Hutton (2013) use Joint model of pre-randomization event counts and multiple post-randomization survival times with cure rates to data for early epilepsy and single seizures, Abeysinghe and Marina (2018) propose Joint Model for Exponential Survival Data and Poisson Count Data to simulate and apply on epilepsy real data and Graeme L. Hickey (2018) apply joint models for longitudinal and competing risks data to epilepsy and compare some alternative joint model for such data.

Many studies have been conducted in various parts of the world and Africa, including Ethiopia that attempted to address many of the issues that arise in connection with the epileptic patients using count regression models for factors of frequency of seizure attacks cross-sectionally and

apply joint model for survival and single count model. However, as far as the researchers' knowledge goes, epilepsy is still a problem worldwide, including in Ethiopia.

There were gaps in previous research on the joint of count longitudinal and survival of time to seizure freedom in epilepsy patients, and their methods failed to identify changes in seizure attack frequency over time and pattern on different covariates, the mean and median time of epileptic patients experiencing seizure freedom related to different covariates, or the duration that speeds up or slows the period of experiencing seizure freedom among epilepsy patients.

Several studies have attempted to use joint models for longitudinal and survival analysis in epilepsy, but a critical research gap persists. Existing methods fail to adequately explore the intricate dynamics of the joint count longitudinal and survival of time to seizure freedom, impeding our ability to discern patterns across different covariates. This knowledge void presents a significant barrier to the development of targeted interventions and personalized treatment strategies for epilepsy patients in HUCSH.

Thus, the goal of this study is to fill this significant gap by using advanced statistical methods to fully investigate the joint count longitudinal and survival characteristics associated with epilepsy. By using this method, we want to determine the covariates that impact seizure occurrences, explore the temporal evolution of seizure patterns, and offer a more nuanced knowledge of the variables determining the time required to become seizure free. Our work aims to address this gap and support the establishment of efficient public health strategies and clinical care optimization for people with epilepsy living in environments with limited resources.

### 1.3 Research Questions

- What are the factors that determine longitudinal evolution of seizure episode of epilepsy patient under the follow up time?
- What are the risk factors determine time to seizure free of epilepsy patient?
- How strong the association between seizure attacks progression and the time to seizure freedom of the epilepsy patients?
- What are the major risk factors that jointly affect the seizure episode and time to seizure freedom of the epilepsy patients?

## 1.4. Objectives of the Study

### 1.4.1. General Objective

The main objective of this study was to identify predictors of seizure attacks progression and time to seizure freedom of epileptic patients using separate and joint model application based on data records from Hawassa University Comprehensive Specialized Hospital

#### 1.4.2. Specific Objectives

- To investigate the change in longitudinal response of seizure attacks over time and the factors that influence change in epilepsy patients separately at HUCSH
- To assess the factors that affect seizure freedom of epilepsy patient separately at HUCSH
- To assess the association level of frequency of seizure attacks and time to seizure freedom.
- To identify determinant risk factor that jointly affects seizure episode and time to seizure freedom of the epilepsy patients under treatment.

## 1.5 Significance of the Study

The results of the study help by provide valuable information for decision-makers in the government of Ethiopia, non-governmental organizations, the Ministry of Health, policy makers, the administration of Hawassa University Comprehensive and Specialized Referral Hospital, enabling them to enhance public awareness regarding factors that contribute to or control epileptic seizures. Furthermore, it will increase awareness in society about the factors that trigger seizures and how to control them. Additionally, the findings can serve as a valuable resource for other researchers interested in conducting studies in similar areas and related topics.

## 1.6 Operational Definitions

- **Adherent**: If the MMAS - 8 score was <6 low adherent, 6-8 medium adherent, 8 high adherent
- Antiepileptic medications are drugs used to control epileptic seizure.
- **Epilepsy**: the word “Epilepsy” does not indicate anything about the cause or severity of the person’s seizures. Having two or more seizures or a tendency to have recurrent seizures, is indicator of epilepsy.

- **Seizure freedom**: An International League Against Epilepsy (ILAE) task force suggested that a patient should be regarded as "seizure-free" in response to a new anti-seizure medication once they have gone without a seizure for at least three times the length of their longest pre-intervention inter-seizure interval in the previous 12 months. (‘‘Rule of Three Principle’’) (Kwanet al., 2009).
- **Seizure**: a seizure is a sudden, uncontrolled electrical disturbance in the brain. It can cause changes in your behavior, movements or feelings, and in levels of consciousness.

## 2. LITERATURE REVIEW

### 2.1General Over View of Epilepsy

The International League Against Epilepsy (ILAE) has developed a practical and clinically useful definition of epilepsy. The ILAE defines epilepsy as a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures, and by the neurobiological, cognitive, psychological, and social consequences of this condition. Epileptic seizures are defined as a transient occurrence of signs and/or symptoms due to abnormal, excessive, or synchronous neuronal activity in the brain. The ILAE emphasizes the importance of considering the patient's medical and social context, as well as the likelihood of future seizures, in the diagnosis and management of epilepsy. This definition is intended to be used by healthcare professionals to facilitate accurate diagnosis, effective treatment, and improved outcomes for people with epilepsy.

According to the World Health Organization, epilepsy affects over 50 million people worldwide, making it one of the most common neurological disorders. The causes of epilepsy vary, and can include genetic factors, head injuries, infections, and brain tumors.

A systematic review and meta-analysis of international studies" by Fiest et al. aimed to determine the prevalence and incidence of epilepsy worldwide through a systematic review and meta-analysis of published studies. The authors analyzed data from 291 studies and found that the global prevalence of epilepsy was 7.60 per 1,000 people, while the global incidence was 67.77 per 100,000 person-years (Fiest, Kirsten M., et al., 2017). The study also found that the prevalence and incidence of epilepsy varied significantly by region, age, and other demographic factors. The authors concluded that this information could help inform public health policies and resource allocation for people with epilepsy around the world.

### 2.2 Related studies on the prognosis factors on the seizure attacks and seizure freedom

Mariam D et al. (2020) examines the treatment outcomes and factors associated with epilepsy in adult patients at Hawassa University Specialized Hospital in Southern Ethiopia. The study found that a majority of patients experienced seizure control with antiepileptic drugs, while a smaller percentage had treatment failure. Factors associated with treatment failure included older age,

being male, and having a longer duration of illness. The authors recommend increased awareness and education about epilepsy and its treatment in the region.

A cross-sectional study was conducted in 2017 to investigate the frequency of seizure attacks and associated factors among patients with epilepsy at the University of Gondar Referral Hospital in North West Ethiopia. The study found that a majority of the participants experienced seizure attacks once or twice a month, with males being more likely to have a higher frequency of seizures than females. Other factors associated with increased seizure frequency included having a family history of epilepsy, longer duration of illness, and not receiving antiepileptic medication. The authors suggest that improving access to antiepileptic medication and promoting awareness about epilepsy could help to reduce the frequency of seizures in this population (Tigistu, Mekdes, et al., 2018).

A remote symptomatic cause, an abnormal EEG, and a history of prior seizures are among the risk variables that increase the risk of a seizure recurrence in patients, according to a research done in 2006 by Kim et al. A more personalized approach to treatment may be required, according to the paper, which also implies that beginning antiepileptic medication after a single seizure may not always lower the chance of recurrence (Kim et al., 2006).

A recent study conducted in Ethiopia presents seizure control and associated factors among pediatric patients with epilepsy at a neurologic outpatient clinic in Ethiopia. The authors found that a significant proportion of the pediatric patients with epilepsy did not achieve optimal seizure control, with factors such as low socioeconomic status, lack of education, and inadequate medication adherence contributing to poor seizure control. They also observed a high prevalence of comorbidities such as cognitive impairment and behavioral disorders among the patients studied (Adal et al., 2021).

Another recent study on epilepsy treatment outcome, adherence to anti-seizure medications and predicting factors at the chronic care facility in Jimma University Medical Center, Jimma, Southwest Ethiopia (Cross-sectional study) found that a significant proportion of patients with epilepsy had poor adherence to their anti-seizure medications, which was associated with poor seizure control. Factors such as lack of education, low socioeconomic status, and forgetfulness were identified as predictors of poor medication adherence. The study also found that patients

who received counseling and support for medication adherence had better seizure control outcomes than those who did not receive such support (Shuma et al., 2022).

A study conducted by Raru et al (2021) Change in the frequency of seizure attacks and associated factors among adult epilepsy patients at Amanuel Mental Specialized Hospital (AMSH) using generalized linear mixed model (GLMM) shows that the frequency of seizure attacks decreased over time in patients who were adherent to their anti-seizure medication. Factors such as low socioeconomic status, lack of education, and inadequate medication adherence were associated with an increased frequency of seizure attacks. The study also found that patients who received counseling and support for medication adherence had better outcomes in terms of seizure control. Overall, the study highlights the importance of medication adherence and support for patients with epilepsy to improve treatment outcomes, particularly in low-resource settings such as Ethiopia.

Recent study identified several factors associated with poor adherence to AEDs, including a lack of understanding about epilepsy and AEDs, lack of social support, and side effects of AEDs. The study also found that patients who had access to health education about epilepsy and AEDs were more likely to adhere to their medication regimen. The authors conclude that poor adherence to AEDs is a significant problem among patients with epilepsy in Ethiopia, and that interventions to improve adherence should focus on increasing patient education and social support. They also suggest that healthcare providers should consider the side effects of AEDs when selecting a medication regimen for their patients (Shegaye S. et al., 2022).

According to the study conducted by (Fanta T, et al., 2015) age, gender, level of education, employment status, and seizure frequency are factors associated with perceived stigma. Patients who had higher levels of education and were employed were less likely to perceive stigma. Patients who had a higher seizure frequency were more likely to perceive stigma. Thus, the study was conducted on a sample of 423 patients with epilepsy, and it was found that 56.6% of them perceived stigma related to their condition. The authors conclude that perceived stigma is a significant problem among patients with epilepsy in Ethiopia, and that interventions to reduce stigma should focus on increasing public awareness and education about epilepsy.

A study conducted in Northwest Ethiopia to assess the seizure control and its associated factors among epileptic patients attending the Neurology Clinic at the University of Gondar hospital involved 422 participants, and data was collected through a structured questionnaire and medical chart review. The study found that 49.3% of participants had controlled seizures, while 50.7% had uncontrolled seizures. Factors such as age, duration of epilepsy, medication adherence, and medication side effects were found to be significantly associated with seizure control (Zena D, et al., 2022)

A systematic review of studies on early predictors of remission in newly diagnosed epilepsy included 31 studies, and the authors identified several factors that were consistently associated with a higher likelihood of remission. These factors included having a normal neurological exam, having a single seizure type, having a normal EEG, and having a shorter duration of epilepsy. The review also identified several factors that were not consistently associated with remission, including age at onset, gender, and family history of epilepsy (Abimbola, Seye, et al., 2014)

A study conducted by (Waja T. et al., 2016) in Addis Ababa, Ethiopia to determine the prevalence of alcohol use disorders and associated factors among people with epilepsy attending Amanuel Mental Specialized Hospital involved 318 participants, and data was collected through a structured questionnaire and medical chart review. The study found that 21.1% of participants had alcohol use disorders, and factors such as male gender, younger age, and lower education level were significantly associated with alcohol use disorders. The study also found that patients with alcohol use disorders were more likely to have uncontrolled seizures and were less likely to adhere to their medication regimen.

Fite et al (2021) used a systematic search of relevant databases to identify articles that reported on the prevalence of stigma among people with epilepsy in Ethiopia. The authors used a meta-analysis to synthesize the results from the selected studies and identify associated factors. The study showed that the pooled prevalence of stigma among people with epilepsy in Ethiopia was 58.3%. Factors associated with stigma included low education level, rural residence, unemployment, longer

## 2.3 Knowledge, attitude and practice of epilepsy

A study conducted by Wubetu et al. (2020) explores the knowledge and attitudes of the public towards epilepsy in Ethiopia. The study was conducted on a sample of 900 participants from three districts, and it was found that 64.2% of participants had poor knowledge about epilepsy, and 56.8% had a negative attitude towards people with epilepsy. The study identified several factors associated with poor knowledge and negative attitudes towards epilepsy, including low educational level, rural residency, and lack of exposure to people with epilepsy. Participants who had higher levels of education and had personal contact with people with epilepsy had better knowledge and more positive attitudes towards the condition. The authors conclude that public knowledge and attitudes towards epilepsy in Ethiopia are inadequate, and that interventions to improve education and awareness about epilepsy are needed. They also suggest that people with epilepsy should be included in public education and awareness campaigns to reduce stigma and discrimination.

A study conducted in Nigeria focuses on the knowledge of and attitude towards epilepsy among women through a cross-sectional survey of 450 women shows the participants had poor knowledge about epilepsy and only a small percentage of the respondents had accurate knowledge of the causes, symptoms, and management of epilepsy. The study also revealed a high level of stigmatization and discrimination towards people with epilepsy, with most participants expressing a desire to keep a distance from individuals with the condition (Fehintola et al., 2019).

A study "Quality of life and associated factors among patients with epilepsy at specialized hospitals, Northwest Ethiopia; 2019" by Minwuyelet et al. used a cross-sectional design and included 416 patients with epilepsy. The study showed that the mean QoL score among patients with epilepsy was low, indicating poor QoL. Factors associated with lower QoL included lower educational status, unemployment, longer duration of epilepsy, frequent seizures, and stigma-related experiences. The authors suggest that interventions aimed at improving QoL for patients with epilepsy should focus on reducing stigma and improving social support, increasing access to education and employment opportunities, and improving epilepsy management, including seizure control and patient education. Overall, the study highlights the importance of addressing

the broader social and economic factors that impact QoL for patients with epilepsy, in addition to providing effective medical treatment.

The prevalence of epilepsy in adults and the challenges in accurately diagnosing and treating seizures is due to communication barriers, atypical seizure presentations, and potential interactions with other comorbid conditions. The study emphasizes the need for a comprehensive approach to epilepsy management, including a thorough assessment of the patient's medical history and neurological exam, appropriate diagnostic testing, and tailored treatment plans that take into account the patient's individual needs and circumstances (Watkins, L. V., et al., 2022).

### 2.4 Joint modeling in medical research

T. Baghfalaki et al. (2020) presents a Bayesian shared parameter model for jointly analyzing longitudinal continuous and binary outcomes. The model incorporates a shared random effect and separate error terms for each outcome. The authors use Markov Chain Monte Carlo (MCMC) methods to estimate the model parameters and perform simulations to evaluate its performance. They also apply the model to a real data example on the joint modeling of HIV viral load and CD4 cell count. The results show that the proposed model outperforms other commonly used models in terms of accuracy and efficiency in estimating the association between the two outcomes.

Karasoy, Duru (2020) presents a joint modeling approach for analyzing longitudinal measurements and survival data in medical studies. Specifically, the proposed method combines a mixed-effects model for the longitudinal measurement and a parametric survival model for the time-to-event outcome. The author applies the joint model to a study on primary biliary cirrhosis, where the longitudinal measurement is the serum bilirubin level and the survival outcome is the time to liver transplantation or death. The results of the joint model reveal that the longitudinal measurement is a significant predictor of the survival outcome, after adjusting for other covariates. The author also compares the joint model with separate models for each outcome and shows that the joint model provides more accurate estimates and better predictive performance. Overall, the joint modeling approach presented in the article can be a useful tool for analyzing complex medical data with both longitudinal and survival components.

A study conducted by Ahmed H. et al. (2017) describes a case study of Bayesian joint modeling of longitudinal and survival data in HIV/AIDS patients at Bale Robe General Hospital in Ethiopia. The authors used Bayesian approach to identify factors associated with disease progression and mortality in HIV/AIDS patients, and to estimate the effect of antiretroviral therapy (ART) on disease progression and survival. The results of their analysis indicated that older age, lower CD4 count, and advanced disease stage were associated with increased risk of mortality, while ART was associated with reduced risk of disease progression and mortality. The authors conclude that Bayesian joint modeling can be a useful tool for analyzing longitudinal and survival data in HIV/AIDS patients, and can provide insights into the effectiveness of ART in improving disease outcomes.

A study by Zhang et al. (2021) describes a Bayesian joint modeling approach to assess the importance of biomarkers in predicting both longitudinal and survival outcomes in the presence of semi-competing risks. The authors used Bayesian approach to analyze data from a study of chronic kidney disease patients, with the aim of identifying biomarkers that are predictive of disease progression and death. The results of their analysis suggested that certain biomarkers, such as albuminuria and estimated glomerular filtration rate, were important predictors of both outcomes, while others were only predictive of disease progression or death. The authors conclude that their approach can be a useful tool for identifying important biomarkers in complex disease processes, and can help inform clinical decision-making by identifying patients at high risk of adverse outcomes.

### 2.5 Previous studies on joint modeling in epilepsy

A study by J.K. Rogers (2013) presents a statistical approach for analyzing data from clinical trials in which participants have a pre-randomization event count (such as the number of seizures experienced prior to the start of the trial) and multiple post-randomization survival times (such as time to recurrence of seizures). The approach incorporates the possibility of a "cure rate," meaning some participants may experience no further events after a certain point in time. The authors apply their approach to data from a clinical trial involving patients with early epilepsy or a single seizure. They compare their method to other commonly used methods and find that it provides more accurate estimates of treatment effects and can detect differences in treatment

effects that other methods may miss. The article concludes that the joint modeling approach can be a valuable tool for analyzing data from clinical trials with complex event histories.

The study conducted by Hickey et al. (2018) compare three different types of joint models for analyzing longitudinal and competing risks data in the context of an epilepsy drug trial. The three models considered were the joint model with shared random effects, the joint model with shared frailties, and the joint model with independent associations. The authors found that the joint model with shared random effects provided the best fit for the data and produced more accurate estimates of treatment effects compared to the other two models. The article provides useful insights for researchers analyzing longitudinal and competing risks data, particularly in the context of clinical trials.

Abeysinghe A. et al (2018) proposes a statistical methodology for analyzing data that involves both survival data and count data. The authors propose a joint model that simultaneously models the time-to-event outcome and the count data, allowing for estimation of the relationship between the two types of data. The model is based on a shared random effect that captures the dependence between the two outcomes, and it allows for the effects of covariates on both outcomes to be estimated. The authors demonstrate the application of the model using a simulated dataset and a real dataset on patients with epilepsy, highlighting the advantages of joint modeling over separate analyses of the two types of data.

A study by Allen and Barnhart (2005) presents a statistical methodology for analyzing data that involves both the frequency of events and the severity of events, with applications to clinical trials. The authors propose a general marginal regression model that simultaneously models the frequency and severity outcomes, allowing for estimation of the relationship between the two types of data. The model is based on a shared random effect that captures the dependence between the two outcomes, and it allows for the effects of covariates on both outcomes to be estimated. The authors demonstrate the application of the model using a simulated dataset and a real dataset from a clinical trial on epilepsy, type II hyperlipoproteinemia and coronary heart patients, highlighting the advantages of joint modeling over separate analyses of the two types of data. The proposed joint model can be used in various fields such as medical and epidemiological research to analyze frequency and severity data jointly.

## 3. DATA and METHODOLOGY

### 3.1. Study Area

The study was conducted at Hawassa University Comprehensive and Specialized Referral Hospital (HUCSH), Hawassa, Ethiopia which is the capital of Sidama Regional State. Hawassa University Comprehensive Specialize Hospital (HUCSH), which started health care delivery services in 2006, is a 400 bedded hospital which provides high quality service at both outpatient and inpatient level for about 20 million populations of the southern regions of the country. The HUCSH also serves as a training center for undergraduate and postgraduate medical students and Health Science trainees [https://hu.edu.et/index.php/about-med/background-med].

### 3.2. Study design and population

A retrospective cohort study design was employed to retrieve relevant information from the medical records of seizure epilepsy patients under follow-up at HUCSH to address the objectives of the study. Epileptic patients age greater than 18 years old were as source of population for this study and the data obtained from HUCSH neurologic Clinic, southern of Ethiopia.

### 3.3 Sample Size Determination

One of the key challenges in joint modeling is determining the appropriate sample size needed to ensure statistical power and accuracy of the model. Chen et al., (2011) present a sample size formula for joint modeling that takes into account the expected effect size, alpha and beta levels, variances and correlations in the longitudinal and survival sub-models. The formula assumes a balanced design and normality of the random effects in the longitudinal sub-model.

The sample size formula for joint modeling is given as follows:

$$d = \frac{(Z_\alpha + Z_\beta)^2}{\sigma_s^2 \alpha^2}$$

And the required total number of subjects (epileptic patients) can be computed as:

$$n = {}^{d}/_{P(event)}$$

where:

α: Association parameter

$$\sigma_s^2 = var(b_{0i}) + \frac{1}{\tau}\left[\frac{2}{\eta^2} - \exp(-\eta\bar{t})(\bar{t}^2 + \frac{2\bar{t}^2}{\eta} + \frac{2}{\eta^2})\right] var(b_{1i}) + \frac{2}{\tau}\left[\frac{1}{\eta} - \exp(-\eta\bar{t})\,\bar{t} + \frac{1}{\eta}\right] cov(b_{0i}, b_{1i})$$

$b_i = (b_{0i}, b_{1i})^T \sim N(0, G)$

$\eta = -\log(0.5)/T_M$

$T_M$: Median event time

$\bar{t}$: Mean follow-up time

$\tau$: Event rate

P: Probability of event

The formula calculates the minimum number of patients needed to detect a certain effect size with a certain level of statistical significance and power, given the variances and correlations in the longitudinal and survival sub models.

The parameters were estimated from pilot survey because no previous joint analyses are available to determine appropriate hypothesized values of association parameters. The objective of the pilot survey in this study was to obtain estimates of the parameters required for sample size calculation. On the basis of the time, cost and labor, pilot survey size n=30 was used.

Table 3.1. The Parameters Used for Sample Size Calculation from the Pilot Survey

| | |
|---|---|
| $T_M$: Median event time | 4 year |
| $\sum\theta_i$ | $\begin{pmatrix}\sigma^2\theta_0 & \sigma\theta_1\theta_0 \\ \sigma\theta_1\theta_0 & \sigma^2\theta_0\end{pmatrix} = \begin{pmatrix}5.83 & 0.63 \\ 0.35 & 3.76\end{pmatrix}$ |
| $\bar{t}$: Mean follow-up time | 3.5 years |
| $\tau$: Event rate | 11% |
| $\eta = -\log(0.5)/T_M$ | 0.188 |

| | |
|---|---|
| $\sigma_s^2$ | 6.49 |
| α | 0.3 |
| P: Probability of event | 0.567 |

Thus, From the total of 1246 registered epileptic patients reported by neurologic clinic at the hospital with in the study period 203 were included in the study.

## 3.4. Data source and method of data collection

The data for this study was collected from Hawassa University Comprehensive and Specialized Referral Hospital (HUCSH), Hawassa, Ethiopia. The epilepsy patients admitted in the hospital from 1st May 2018 up to 1st May 2023 was a source of population for this study. The longitudinal and survival data was collected from primary data collection method using well-structured questionnaires and secondary data collection method using patient chart. The longitudinal episodes of seizure is measured by counting the seizure per year. The survival endpoint of interest was the time to seizure freedom of the epilepsy patients. Both primary and secondary data was collected by a trained nurse and statistician.

### 3.5 Eligibility criteria

**i. Inclusion criteria:**

- Patient's chart with complete records.
- Patients who were on AEDs treatment for at least the last 12 months (a year) consecutively.
- Epilepsy is more likely to develop in older adults because some risk factors for epilepsy are more common in older adults, such as: strokes, head injuries from falls, diseases that affect brain function (such as Alzheimer's disease), brain tumors (Brodie MJ et al., 2009). Therefore, patients' age ≥18 years included in the current study
- Voluntary epileptic patients for interview
- A history of two or more clinically definite unprovoked seizures occurring at least 24 hours apart

**ii. Exclusion criteria:**

- Patients below age 18

- Epilepsy patients who were followed for less than one year and
- Patients who were registered before 1st May 2018

### 3.5. Variables in the Study

#### 3.5.1. Response Variables

For this study two response variables were considered:

**i. longitudinal response**

Frequency of seizure attacks which is the number of uncontrolled involuntary shaking involving only one part of the brain or involving the entire brain. The frequency of seizure is measured by counting the seizure per year.

**ii. Survival response**

The survival outcome of patients was time to seizure freedom, i.e. the time start of the study until the patients became first seizure free period for the last 12 months.

$$status = \begin{cases} 1 & if, event\ occur = siuzer\ free \\ 0 & if, censore \end{cases}$$

Since there are censored observations the coding of the status variable is such that 1 indicates uncensored observation (for those who became seizure free for 12 months in their follow-ups period) and 0 for censored observations (for those who was not seizure free at the time of the survey).

#### 3.5.2. Independent variables

From different literature the researcher found the possible independent variables for this study that means socio-demographic variables and clinical factors.

- Socio-demographic variables and their code
  - Sex(0=Female,1=Male)
  - Age at base line
  - Residence(0=Urban,1=Rural)
  - Religion (0=Protestant, 1=Orthodox, 2=Muslim, Others=3)

- Marital status(0=Single1=Married,2=Divorced,3=widowed)
- Education level (0=illiterate,1=religious school,2=primary, 3=secondary, 4=tertiary)
- Occupation (0=Gov't employee ,1=Student,2=Farmer,3=Skilled labour,4=Others)

➢ Clinical variables

- Age at seizure onset
- Type of AED (0= Carbamazepine. 1= Phenobarbitone,2 = Phenytoin, 3= Sodium Valporate, 4= more treatment)
- Exercise(0=No,1=yes)
- Family history of epilepsy (0=No,1=Yes)
- Co-morbidity(0=No1=Yes)
- Sleep deprivation(0=No,1=Yes)
- Traditional drug taking(0=No,1=Yes)
- Epilepsy is controlled by modern drug (0=No,1=Yes)
- Stress(0=No,1=Yes)
- Alcohol consumption (0=No,1=Yes)
- Traumatic brain (head injury (0=No, 1=Yes))
- Medication source (0=Free,1=Payment)
- Chronic disease (0=DM,1=Hyepertensive,2=HIV,3=Others,4=N0)
- Adherence Level (0=Low,1=Medium,2=High)

### 3.6. Methods of Data Analysis

In this study, various statistical models were considered, including linear mixed-effects models for count longitudinal data, survival models for time-to-event data and joint model for longitudinal and survival data. Longitudinal measurements of seizure attacks and time to seizure freedom in epileptic seizure patients were analyzed using subject-specific random effects with the INLA package in R (software version 4.0.0), and data entry was performed using SPSS version 25.0. A statistical decision was made at a 5% significance level.

#### 3.6.1. Descriptive Statistics and Data exploration

The frequency table (for data exploration), individual profiles plot, and mean profile plot (for the longitudinal dataset), as well as the Kaplan-Meier curve (for the survival dataset), were all considered.

### 3.6.2. Separate Analysis of Longitudinal Data

#### *3.6.2.1 Separate Analysis of Longitudinal Data*

With longitudinal data, the same set of individuals are observed repeatedly across a finite number of time points at equally or randomly spaced intervals. Longitudinal data are therefore the results of repeated measurements at a fixed number of time points with predefined designs on time scale, time interval, and other relevant factors (Xian, 2015).

For every individual, numerous observations are gathered over time in longitudinal research. As a result, longitudinal studies are described as those in which the outcome variable is assessed repeatedly, i.e., in the same person on many times. Because the repeated observations of each individual are correlated, it is important to use particular statistical approaches in longitudinal studies where observations of one individual over time are not independent of one another (Jos W. R. Twisk, 2003).

**Generalized Linear Mixed Model**

Two sources of variation are taken into account for longitudinal data. These are the within-subject variations that occur during measurements of each subject (epileptic patients), as well as the between-subject variations that occur during measurements of various subjects. For this study, modeling between subject variations helps us understand distinctions between subjects, such as between epileptic patients, while modeling within subject variation aids in the analysis of changes over time. Therefore, for this study generalized linear mixed effects model (GLMEM) is used to conduct separate analysis of frequency count for seizure longitudinal event. Here are the equations for mixed effect Poisson, negative binomial and zero-inflated regression models for longitudinal data:

**Generalized Linear Mixed Poisson Regression Model:**

The Poisson distribution is often used to model count data, such as the number of events that occur within a fixed interval of time. A generalized linear mixed Poisson regression model can be used to model the relationship between a count outcome variable Y and one or more predictor variables X, while accounting for the correlation among observations within the same subject (or cluster). The mixed effect Poisson regression model can be written as:

$$\log\left(E\left(Y_{ij}\right)\right) = \beta_0 + \beta_i X_{ij} + \tau_i t_{ij} + b_i + \varepsilon_i$$

where:

- $Y_{ij}$ is the count outcome variable for the i-th subject at the j-th time point
- $X_{ij}$ is the vector of predictor variables for the i$^{th}$ subject at the j$^{th}$ time point
- $\beta_0$ and $\beta_i$ are the intercept and slope coefficients, respectively, for the fixed effects
- $\tau_i \; and \; t_{ij}$ are coefficients and time variable respectively
- $b_i$ is the random effect for the i$^{th}$ subject, assumed to be normally distributed with mean 0 and variance $\sigma_b^2$

**Generalized Linear Mixed Negative Binomial Regression Model:**

The negative binomial distribution is often used to model overdispersed count data, where the variance exceeds the mean. A mixed effect negative binomial regression model can be used to model the relationship between a count outcome variable Y and one or more predictor variables X, while accounting for the correlation among observations within the same subject (or cluster) and the overdispersion of the count data. The generalized linear mixed negative binomial regression model can be written as:

$$\log\left(E\left(Y_{ij}\right)\right) = \beta_0 + \beta_i X_{ij} + b_i + \varepsilon_i$$

$$\log\left(Var\left(Y_{ij}\right)\right) = \log\left(E\left(Y_{ij}\right)\right) + \sigma^2$$

$$\log(\varepsilon_i) = c_0 + c_1 + Z_i$$

where:

- $Y_{ij}$ is the count outcome variable for the i-th subject at the j-th time point
- $X_{ij}$ is the vector of predictor variables for the i-th subject at the j-th time point
- $\beta_0$ and $\beta_i$ are the intercept and slope coefficients, respectively, for the fixed effects
- $b_i$ is the random effect for the i-th subject, assumed to be normally distributed with mean 0 and variance $\sigma_b^2$

- $\sigma^2$ is the overdispersion parameter, which accounts for the excess variance in the count data relative to the Poisson distribution.
- $Z_i$ are subject-specific covariates used to model the over-dispersion in the count data
- $c_0$ *and* $c_1$ are the fixed effects for the intercept and subject-specific covariate respectively

**Zero-inflated model**

**The zero-inflated Poisson model**

In a longitudinal setting, suppose that $Y_{ij}$ is the response such as the symptom count for subject *i* at time $t_{ij}$ (i=1,…,n, j=1,…,$m_i$ ) and it is from a finite mixture:

$$Y_{ij} \sim \begin{cases} 0, & with\ probability\ \pi_{ij} \\ Poisson(\lambda_{ij}), & with\ probability\ 1-\pi_{ij} \end{cases}$$

The probability distribution is thus written as

$$P(Y_{ij}=0) = \pi_{ij} + (1-\pi_{ij})e^{-\lambda_{ij}}$$

$$P(Y_{ij}=y_{ij}) = (1-\pi_{ij})\frac{e^{-\lambda_{ij}}\lambda_{ij}^{y_{ij}}}{y_{ij}!}, y_{ij} = 1,2,\ldots$$

**The zero-inflated Negative Binomial model**

$$Y_{ij} \sim \begin{cases} 0, & with\ probability\ \pi_{ij} \\ NB(\lambda_{ij}, a), & with\ probability\ 1-\pi_{ij} \end{cases}$$

The probability distribution is thus written as

$$P(Y_{ij}=0) = \pi_{ij} + (1-\pi_{ij})\left(\frac{a^{-1}}{a^{-1}+\lambda_{ij}}\right)^{a^{-1}}$$

$$P(Y_{ij}=y_{ij}) = (1-\pi_{ij})\left[\frac{\Gamma(y_{ij}+a^{-1})}{\Gamma a^{-1} y_{ij}!}\left(\frac{a^{-1}}{a^{-1}+\lambda_{ij}}\right)^{a^{-1}}\left(\frac{\lambda_{ij}}{a^{-1}+\lambda_{ij}}\right)^{y_{ij}}\right], y_{ij} = 1,2,\ldots$$

**Model Selection**

Akaike information criteria (AIC), Bayesian information criteria (BIC), and Deviance were applied for model selection after fitting the appropriate model for the data by assuming random

intercept only model, the random slope only model, and random intercept and random slope model.

### 3.6.2.2 Separate Analysis of Survival Data

The time until an event occurs is the outcome variable of interest for a family of statistical techniques for data analysis known as survival analysis. Time is defined as the number of months from the start of a person's follow-up until an event (seizure free) happens. The duration from a given origin to the occurrence of an event of interest is measured by survival data, also known as time to event data.

**Censoring**

Survival analysis is most important statistical method when there are censoring data as opposed to the use of different statistical methods. Censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly. May occur when a person does not experience the event before the study ends, a person is lost to follow-up during the study period and a person withdraws from the study because of different reasons, example death (if death is not the event of interest). For this study if epileptic patients not experience seizure free throughout the study period they were considered as censored (right censoring).

In survival analysis there are two functions of central interest namely the survivor function and the hazard function.

**i. Survivor Function S(t)**

Let ***T*** be the random variable for a person's survival time and let ***t*** be any specific value of interest for the random variable ***T***, then the cumulative distribution function F(t) and the survivor function, denoted by $S(t)$ is given by

$$F(t) = p(T \leq t) = \int_0^t f(u)du, \;\; t \geq 0$$

Where, $F(t)$ the probability that a patient gets seizure freedom before time t.

$$S(t) = p(T > t) = 1 - F(t) = 1 - \int_0^t f(u)du, \;\; t \geq 0$$

- The cumulative distribution function F(t) gives the probability that a subject selected at random will have a survival time less than some stated value t.
- The survivor function S(t) gives the probability that a person survives longer than some specified $t_i$: that is, S(t) gives the probability that the random variable T exceeds the specified time t.

**The properties of the survival function are:**

- they are non-increasing; that is, they head downward as t increases;
- at time t =0, S(t) =S(0) =1; that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time 0 is one;
- at time t =∞, S(t)=S(∞)=0; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually fall to zero.

**ii. Hazard Function h(t)**

The hazard function is widely used to express the risk of hazard of recover at time t. It is obtained from the probability that an individual dies at time t, given that the individual has survived up to time t. in this study the event time to seizure freedom hazard rate is the experiencing rate of seizure freedom. For the given T and t the hazard function, denoted by h(t), is given by

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

The hazard function h(t) gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t.

**The properties of the hazard function are:**

- it is always nonnegative, that is, equal to or greater than zero;
- it has no upper bound
- it may be used to identify a specific model form, such as an exponential, a Weibull, or a lognormal curve that fits one's data

**Estimation of the survival function**

**Non-parametric methods for survival data analysis**

In survival analysis, it is always a good idea to present numerical or graphical summaries of the survival times for the individuals. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time.

The Kaplan-Meier is the most widely used to estimate the survival and hazard functions of descriptive survival analysis. These methods are said to be non-parametric or distribution free since they do not require specific assumption to be made about the underlying distribution of the survival time.

**Kaplan-Meier (KM)**

The standard estimator of the survival function, proposed by Kaplan and Meier (1958), is called the Product-Limit estimator. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. We use the observed data to estimate the conditional probability of confirmed survival at each observed survival time and then multiply them to obtain an estimate of the overall survival function.

Suppose $t_1, t_2, \ldots, t_n$ be the observed survival times for *n* subject and consider that *m* subjects experienced the event, $t_{(1)} \leq, t_{(2)} \leq \ldots \leq t_{(m)}$, $m \leq n$ being the *m* distinct ordered survival times. Then the KM estimator of the survivor function at time *t* is given by

$$\hat{S}_{KM}(t) = \begin{cases} 1, & if\ t \leq t_{(1)} \\ \prod_{t(i) \leq t} \left( \frac{n_i - d_i}{n_i} \right), & if\ t \geq t_{(1)} \end{cases}$$

Where,

$d_i$ = The number of epileptic patients who experience the event (seizure freedom) at time $t_i$

$n_i$ = The number of patients who have not yet experienced the event (seizure freedom) or the total number of individuals who experienced the event before at time $t_i$

**Log-Rank Test (Comparison of Survivorship Functions)**

First descriptive statistics which provide a description of the overall survival experience is needed, then one is expected to proceed with a comparison of the survivorship experience of subgroups in the data. These groups might be defined by the values of a covariate which are thought to be related to survival times. When comparing groups of subjects, it is always preferable to begin with a graphical display of the data in each group.

The figure in general shows if the pattern of one survivorship function lying above another which means the group defined by the upper curve lived longer, or had a more favorable survival experience, than the group defined by the lower curve. Now the statistical question is whether the observed difference seen in the figure is significant. The log-rank test is the most commonly used statistical test for comparing the survival distributions of two or more groups.

Peto and Peto (1972) proposes a log-rank test, to compare the whole survival functions of different k groups, for k ≥ 2, of which the hazard ratio comparing two groups is constant over time (proportional hazard) given hypothesis as $H_0: \hat{S}_1(t) = \hat{S}_2(t) = \cdots = \hat{S}_k(t)$; means there is no difference between the survival time of different groups at a given time point. $H_0$: At least the survival function one group is different from the others; where 0 < t < T.

The test statistics is given as

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{Var(O_i - E_i)} \approx \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

which has approximately the chi-square distribution with k−1 degrees of freedom. Where $O_i$ is the observed number of deaths in group *i* and $E_i$ is the expected deaths in the same group. A large $\chi^2$ value would lead to the rejection of the null hypothesis in favor of the alternative hypothesis.

**Survival model**

The non-parametric methods are not useful for controlling the covariates and it requires categorical predictors. Therefore, the multivariable approaches are used when we have several prognostic variables. In survival analysis, we are mainly interested in the risk or hazard of death/failure at any time after the starting point of the study for this survival time is the experiencing of seizure freedom. A variety of models and methods have been developed for doing this sort of survival analysis using either parametric or semi-parametric approaches.

**Cox proportional hazard model**

One of the most popular types of regression models used in survival analysis is the Cox proportional hazard model introduced by cox (Cox, 1972). It is the most popular because it is simple; it can easily accommodate right censoring: that is, the presence of subjects in the data set who have not yet experienced a failure by the end of the study period, and it is used to relate several risk factors or exposures, considered simultaneously, to survival time. It is essentially a regression model commonly used statistical in medical research for investigating the association between the survival time of patients and one or more predictor variables and to obtain an estimate of the hazard function itself for an individual.

The proportional hazards model proposed by Cox (1972) has the following forms:

$$\lambda(t|x) = \lambda_0(t)\exp\left(\sum_{i=1}^{p} \beta_i x_i\right) = \lambda_0(t)\exp\left(x'\beta\right)$$

where

- $(t|x)$ denote the hazard function at time $t$ for the $i$th subject, $i = 1,2, \cdots, n$, with covariate values $x1, x2, \cdots, xp$.
- $\lambda o(t)$ is the baseline hazard function and represents the hazard when all of the predictors (or independent variables) $x1, x2, \cdots, xp$ included in the model equal to zero (i.e., “the reference group")..
- $x$ is a $p \times 1$ vector of explanatory variables
- $\beta$ is a $p \times 1$ vector of unknown regression coefficients.

**Hazard ratio (HR)**

Given two sets of covariates $x_1$ and $x_0$, then the ratio of their hazards at time $t$ is given by:

$$\boldsymbol{H(t) = \frac{\lambda(t, x_1)}{\lambda(t, x_0)} = \frac{\lambda_0(t)exp(x_1'\beta)}{\lambda_0(t)exp(x_0'\beta)} = exp(x_1 - x_0)'\beta, \forall t \geq 0}$$

This model is referred to as the Cox model, the Cox proportional hazards model or simply the proportional hazards model. The term proportional hazards refers to the fact that, the hazard functions are multiplicatively related (i.e., their ratio is constant over survival time).Therefore, in this study the Cox regression model extends survival analysis methods to assess simultaneously the effect of several risk factors on survival time epilepsy patient.

**Assumption of Proportional Hazard Model**

Gillespie (2006) formulates the assumptions of cox regression model as follows.

- Cox regression analysis assumes that the censoring in the data is random or non-informative. This means that patients following up at time "t"(dropping the patients who had an event and have been censored) would be a random sample from the entire study population.
- The baseline hazard function, $\lambda_0(t)$ in the above Cox model, can take any form, but it cannot be negative.

**Parametric survival Models**

The parametric models in survival analysis are used to know how covariates influence the response variable. The parametric models that used to fit survival model are Exponential, Weibull and Lognormal.

**i) Exponential Regression Model**

If the survival time of hypertensive patients has an exponential distribution with fixed hazard rate $\lambda$, then the probability density function of time (the probability of dying at any time) is

$$f(t) = \lambda \exp\{-\lambda t\}$$

The survivor function is

$$S(t) = \exp\{-\lambda t\} \text{ and}$$

Hazard rate is

$$h(t) = \lambda \text{ does not depend on time.}$$

**ii) Weibull Regression Model**

For positive parameters $\lambda, \alpha$; the density function of time is given as:

$$\boldsymbol{f(t) = \lambda\alpha t^{\alpha-1} exp\{-\lambda t^{\alpha}\};} \quad t > 0\ ,$$

And, the survival function (S(t)) is given by:

$$\boldsymbol{S(t) = P(T > t) = \exp\{-\lambda t^{\alpha}\};} \qquad t > 0$$

And, the hazard function is equal to:

$$\boldsymbol{h(t) = \frac{f(t)}{s(t)} = \lambda\alpha t^{\alpha-1}\ ;} \qquad t > 0$$

If $\boldsymbol{\alpha = 1}$ this model reduces to the exponential and has constant risk over time. If $\boldsymbol{\alpha} > 1$, then the risk increases over time. If $\boldsymbol{\alpha} < 1$, then the risk decreases over time. In fact, taking logs in the expression for the hazard function, we see that the log of the Weibull risk is a linear function of log time with slope $\boldsymbol{\alpha - 1.}$

**iii) Lognormal Regression Model**

The log-normal model may take censored time dependent variable that allows the hazard rate to increase and decrease. The log-normal model assumes that $\varepsilon \sim N\ (0, 1)$. The hazard function of $T$ when $\beta_i = 0$ is $h(t) = \frac{\phi(t)\left(\frac{\log(t)}{\sigma}\right)}{\left[1-\phi(t)\left(\frac{\log(t)}{\sigma}\right)\right]\sigma t}$

Where $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right)$ is a probability density function and obviously, we no longer have a proportional hazard model. The survival function $S(t/X)$ at any covariate x can be expressed as: $S(t/X) = \Phi\left(\beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \cdots + \beta_k^* x_k - \frac{1}{\sigma}\log(t)\right)$

Where $\beta_j^* = \frac{\beta_j}{\sigma}$

### 3.7. Joint Modeling of Longitudinal and Survival Data

Joint models of longitudinal and time-to-event data are statistical models that allow for the analysis of both longitudinal and survival data simultaneously. In these models, the longitudinal data is used to model the trajectory of a continuous or discrete outcome over time, while the survival data is used to model the time until an event of interest occurs, such as death, relapse, or disease onset. The joint modeling approach allows for the incorporation of the longitudinal data in the survival analysis, and vice versa, which can lead to more efficient and accurate estimation of the association between the two types of data (Rizopoulos D., 2012).

One commonly used type of joint model is the joint model of longitudinal and survival data proposed by Henderson and colleagues in 2000 (Henderson, R et al., 2000). This model combines a linear mixed-effects model for the longitudinal data and a Cox proportional hazards model for the time-to-event data, and allows for the estimation of both the longitudinal trajectory and the hazard function for the time-to-event outcome.

### 3.7.3 Count Sub-Longitudinal Model

Assuming that the count data for a set of individuals measured at different time points, and we want to model the relationship between the counts and some predictor variables. We can use a Poisson regression model with random effects to account for the correlation within individuals over time. The model can be written as:

$$y_{ij} \sim Poisson(\mu_{it})$$

$$\log(\mu_{ij}) = \beta_0 + \beta_1 time_{ij} + X_{ij}\beta_{2i} + Z_i + b_i + \varepsilon_{ij}$$

where

- $y_{ij}$ is the count for the ith individual at time t,
- $\mu_{ij}$ is the corresponding mean count,
- $\beta_0$ is the intercept, $\beta_1$ is the regression coefficient for time, $\beta_{2i}$ is the regression coefficient for the group variable (which is binary), and
- $time_{ij}$ is the time at the jth occasion for the ith subject
- $X_{ij}$ is the matrix of predictor variables for the ith individual at time t, beta is the vector of fixed effects coefficients,
- $Z_i$ is the matrix of random effects predictors for the ith individual, and
- $b_i$ is the vector of random effects for the ith individual.
- $\varepsilon_{ij}$ is the residual error

The negative binomial sub-longitudinal model is similar to the Poisson model, but it accounts for overdispersion, which occurs when the variance is greater than the mean. The model assumes that the count data follows a negative binomial distribution with a mean that varies between subjects. Let $y_{ij}$ be the count for the ith subject at the jth occasion, then the model is:

$$y_{ij} \sim NB(\mu_{it})$$

$$\log(\mu_{ij}) = \beta_0 + \beta_1 time_{ij} + X_{ij}\beta_{2i} + Z_i + b_i + \varepsilon_{ij}$$

$$\log(\mu_{ii}) = \mu + \alpha * \log(\mu_{ij})$$

where

- $\beta_0$ is the intercept, $\beta_1$ is the regression coefficient for time, $\beta_{2i}$ is the regression coefficient
- $time_{ij}$ is the time at the jth occasion for the ith subject
- μ is the intercept for the log of the variance, and
- α is the regression coefficient for the log of the mean.
- $\varepsilon_{ij}$ is the residual error

**Zero-inflated model**

- **ZIP longitudinal sub model**

Assuming independence between subjects, let $Y_{ij}$ denote the longitudinal response (episode of seizure) for subject $i = 1, .., n$ at time j for j $= 1, \ldots, J_i$ . The distribution of $Y_{ij}$ is

$$Y_{ij} \sim \begin{cases} 0, & with\ probability\ \pi_{ij} \\ Poisson(\lambda_{ij}), & with\ probability\ 1 - \pi_{ij} \end{cases}$$

where $\pi_{ij}$ denotes the probability of the observation arising from the degenerate distribution at zero and $\lambda_{ij}$ represents the mean of the Poisson distribution. The probability distribution function is

$$P(Y_{ij} = 0) = \pi_{ij} + (1 - \pi_{ij})e^{-\lambda_{ij}}$$

$$P(Y_{ij} = y_{ij}) = (1 - \pi_{ij})\frac{e^{-\lambda_{ij}}\lambda_{ij}^{y_{ij}}}{y_{ij}!}, y_{ij} = 1,2, \ldots$$

- **ZINB longitudinal sub model**

The flexible ZINB model is used to adjust for dispersion in an otherwise ZIP model with an additional dispersion parameter $a$, which is positive for overdispersed and negative for

underdispersed data. Let $Y_{ij}$ denote the episode of seizure for subjects for subject $i = 1,..,n$ at time j for j $= 1,\ldots,J_i$. Then $Y_{ij}$ has the following distribution

$$Y_{ij} \sim \begin{cases} 0, & with\ probability\ \pi_{ij} \\ NB(\lambda_{ij}, a), & with\ probability\ 1-\pi_{ij} \end{cases}$$

The probability distribution is thus written as

$$P(Y_{ij}=0) = \pi_{ij} + (1-\pi_{ij})\left(\frac{a^{-1}}{a^{-1}+\lambda_{ij}}\right)^{a^{-1}}$$

$$P(Y_{ij}=y_{ij}) = (1-\pi_{ij})\left[\frac{\Gamma(y_{ij}+a^{-1})}{\Gamma a^{-1} y_{ij}!}\left(\frac{a^{-1}}{a^{-1}+\lambda_{ij}}\right)^{a^{-1}}\left(\frac{\lambda_{ij}}{a^{-1}+\lambda_{ij}}\right)^{y_{ij}}\right], y_{ij} = 1,2,\ldots$$

### 3.7.2 Sub-Survival Model

Assuming that we have survival data for a set of individuals, and we want to model the relationship between the survival times and some predictor variables. We can use a Cox proportional hazards model with random effects to account for the correlation within individuals. The model can be written as:

$$h_i(t) = h_0(t) * \exp\left(X_i(t)\beta + Z_i * b_i * \varepsilon_{ij}\right)$$

where $h_i(t)$ is the hazard rate for the ith individual at time t,

- $h_0(t)$ is the baseline hazard which can be Exponential, Weibull or Lognormal distribution,
- $X_i(t)$is the matrix of predictor variables for the ith individual at time t, beta is the vector of fixed effects coefficients,
- $Z_i$ is the matrix of random effects predictors for the i[th] individual, and
- $b_i$ is the vector of random effects for the ith individual
- $\varepsilon_{ij}$ is the residual error

### 3.7.4 The Joint Model Structure

Joint a generalized linear mixed-effects model for the longitudinal count data and a Cox proportional hazards model for the survival data. The joint model can be expressed mathematically as follows:

$$h(t|x_i, b_i) = h_0(t)\exp\left(\gamma' z_i + \lambda y_i + \alpha b_i\right)$$

$$\log\left(E\left(Y_{ij}|b_i\right)\right) = \mu_{ij} + \eta_i$$

where:

- $h(t|x_i, b_i)$ represents the hazard function for the survival data at time t for individual i with covariate vector $x_i$ and random effects $b_i$.
- $h_0(t)$ is the baseline hazard function.
- γ is a vector of fixed effects coefficients for the survival model covariates.
- $z_i$ is a vector of covariates for the survival model.
- λ is the coefficient for the association between the longitudinal count outcome and the survival outcome.
- $Y_{ij}$ is the longitudinal count outcome for individual i.
- α is a scalar association parameter between the random effects in the survival and longitudinal models.
- $\mu_{ij}$ is the mean value for the longitudinal count outcome for individual i at time t.
- $\eta_i$ is the vector of random effects for the longitudinal count model.
- $E\left(Y_{ij}|b_i\right)$is the expected value of the longitudinal count outcome for individual i at time t given their random effects $b_i$.

### 3.7.5 Method of Estimation for Joint Model

In the statistical field of parameter estimation, the parameters of a distribution are estimated using sample data. The term "estimators" refers to the methods used for parameter estimation. A two-stage technique, a Bayesian method (also known as the Markov chain Monte Carlo method), and a maximum likelihood approach are the primary types of estimating methods used in joint models of longitudinal and survival outcomes. The joint model literature uses the maximum likelihood method, which is one of these techniques, potentially the most commonly.

When the dimension of the random effects is high, the computational complexity of inference for joint models becomes a significant obstacle. The longitudinal measurement and time to event procedures share certain parts of their multivariate Gaussian process in our joint modeling. It may be advantageous to use a simultaneous approach of inference based on the joint likelihood

of longitudinal measurements and times to event, but the computing challenges can be significant. The complexity of these issues can be lowered using a Bayesian approach.

The likelihood, prior, and posterior distributions are the key components of Bayesian statistical inference. In the context of a joint count longitudinal and survival model, the likelihood, prior, and posterior distributions for the model parameters can be expressed as follows:

### 3.7.5.1 Joint Model Likelihood

The likelihood function for the joint count longitudinal and survival model can be written as the product of the individual likelihoods for the longitudinal and survival components of the model:

The likelihood function is given by:

$$L(\theta|y,t) = \prod_{i=1}^{n}[f(y_i,t_i|\theta)]^{c_i}\exp\{-\int_0^{t_i} h(u|\theta)du$$

where:

- θ is the vector of parameters to be estimated
- y is the vector of longitudinal count data
- t is the vector of survival times
- $c_i$ is the censoring indicator for the $i^{th}$ subject
- $f(y_i,t_i|\theta)$ is the joint density of the longitudinal count data and the survival time
- $h(u|\theta)$ is the hazard function for the survival time

### 3.7.5.2 Prior Distributions

The prior distribution for the model parameters can be specified using a suitable prior distribution for each parameter.

The prior distribution of the parameters is given by:

$$p(\theta) = \prod_{j=1}^{p} p(\theta_j)$$

where:

- p is the number of parameters to be estimated
- $p(\theta_j)$ is the prior distribution of the $j^{th}$ parameter

Noninformative joint prior distribution of the parameters are considered: β's and θ's are normally distributed with mean zero and large variance 1000, association parameters α are assumed to have normal distribution with mean zero and variance 1000.

### 3.7.5.3 Posterior Distributions

The posterior distribution for the model parameters can be computed using Bayes' theorem by multiplying the likelihood and the prior distribution:

$$p(\theta|y,t) \propto L(\theta|y,t) * p(\theta)$$

The joint posterior distribution of the parameters and the random effects is given by:

$$p(\theta,u|y,t) \propto L(\theta,u|y,t) * p(\theta) * p(u)$$

where:

- u is the vector of random effects
- p(u) is the prior distribution of the random effects

### Integrated Nested Laplace Approximation (INLA)

Integrated Nested Laplace Approximations (INLA) were introduced by Rue, Martino and Chopin (2009) as a tool to do approximate Bayesian inference in latent Gaussian models (LGMs). The class of LGMs covers a large part of models used today among many models' survival model, generalized linear mixed models (GLMMs) and joint longitudinal-survival models are example, and the INLA approach has been shown to be very accurate and extremely fast in most cases. Software is provided through the R-INLA package, see http://www.r-inla.org.

Latent Gaussian models are hierarchical models which assume an *n*-dimensional Gaussian field **x** = $\{x_i : i \in V\}$ to be point-wise observed through n-dimensional conditional independent data **y**. Both the covariance matrix of the Gaussian field **x** and the likelihood model for $y_i|\mathbf{x}$ can be controlled by some unknown hyperparameters θ. In addition, some linear constraints of the form **Ax = e**, where the matrix A has rank *k*, may be imposed. The posterior then reads:

$$\pi(x,\theta|y) \propto \pi(\theta)\pi(x|\theta)\prod_{i\in I}\pi(y_i|x_i,\theta)\,\tilde{\pi}$$

The INLA approach provides a recipe for fast Bayesian inference using accurate approximations of the marginal posterior density for the hyperparameters $\tilde{\pi}$ (θ |y) and for the full conditional posterior marginal densities for the latent variables $\tilde{\pi}$ ($x_i$|θ , y), i = 1,..., n.

Posterior marginals for the latent variables $\tilde{\pi}$ ($x_i$|y) are then computed via numerical integration such as:

$$\tilde{\pi}(x_i|y) = \int \tilde{\pi}(x_i|\theta,y)\,\tilde{\pi}(\theta|y)d\theta$$

$$\approx \sum_{k=1}^{K} \tilde{\pi}(x_i|\theta_k,y)\,\tilde{\pi}(\theta_k|y)\,\Delta_k$$

INLA presents two main advantages over MCMC techniques. The first and most outstanding is computational. Using INLA results are obtained in seconds and minutes even for models with a huge dimensional latent field, while a well build MCMC algorithm would take hours or even days. This is also due to the fact that INLA is naturally parallelized, thus making it possible to exploit the new trend of having multi-core processors. The second, and not less important advantage, is that INLA treats latent Gaussian models in a unified way, thus allowing greater automation of the inference process. The core of the computational machinery, automatically adapts to any kind of latent field so that, from the computational point of view, it does not matter if we deal with, for example, spatial or temporal models. In practice INLA can be used almost as a black box to analyses latent Gaussian models.

**Diagnostic for all model**

Several diagnostic tests have been devised to track the algorithm's convergence. Among several ways of a test of convergence, the most popular and straightforward convergence assessment methods had been used. The following methods were considered for this study

**Density plot**: This is also another method or technique which can be taken for checking convergence in the Bayesian analysis. The idea is that the Markov chain has attained its posterior

distribution when the density plots of the independent variable coefficients are normally distributed (Merkle et al. 2005).

# 4. Results and Discussions

4.1. Descriptive Results

The baseline socio-demographic and clinical characteristics of epileptic patients included in the analysis presented in Table 4.1. The study has included 203 epileptic patients, who started anti-epileptic drugs (AEDs) at Hawassa University comprehensive specialized hospital between 1st May 2018 up to 1st May 2023.

From the longitudinal response variable seizure attacks which was calculated every 12 month and the time to event response variable was time to seizure freedom patients. To say epileptic patients be seizure free they must be no seizure for the last 12 months on their follow-up times. And an individual was censored if he had no free of seizure for a year until the study ended. Hence, the time-to- the first Seizure freedom in months was the total time of follow ups until patient being seizure free for the previous one year and total follow up times for censored patients. Thus, from the total of 203, 163(80.3%) patients were seizure free while the rest 40(19.7%) were censored patients.

Table 4.1: Summary of sociodemographic variables

| Variables | Category | Frequency | Percentage | Status (Baseline) | |
|---|---|---|---|---|---|
| | | | | Seizures free | Censored |
| Sex | Male | 119 | 58.6 | 118(58.1%) | 1(0.5%) |
| | Female | 84 | 41.4 | 21(10.3%) | 63(31.0%) |
| Occupation | farmer | 33 | 16.3 | 1(0.5%) | 32(15.8%) |
| | government employee | 132 | 65.0 | 18(8.9%) | 114(56.2%) |
| | no work | 5 | 2.5 | - | 5(2.5%) |
| | skill labor | 13 | 6.4 | 2(1.0%) | 11(5.4%) |
| | student | 20 | 9.9 | 1(0.5%) | 19(9.4%) |
| Education level | illiterate | 39 | 19.2 | 3(1.5%) | 36(17.7%) |
| | primary | 15 | 7.4 | 1(0.5%) | 14(6.9%) |
| | secondary | 22 | 10.8 | 1(0.5%) | 21(10.3%) |
| | tertiary | 127 | 62.6 | 17(8.4%) | 110(54.2%) |

| | | | | | |
|---|---|---|---|---|---|
| Resident | rural | 63 | 31.0 | 7(3.4%) | 56(27.6%) |
| | urban | 140 | 69.0 | 15(7.4%) | 125(61.6%) |
| Marital status | divorced | 14 | 6.9 | - | 14(6.9%) |
| | married | 136 | 67.0 | 17(8.4%) | 119(58.6%) |
| | single | 37 | 18.2 | 4(2.0%) | 33(16.3%) |
| | widowed | 15 | 7.4 | 14(6.9%) | 1(0.5%) |
| Living status | alone | 30 | 14.8 | 2(1.0%) | 28(13.8%) |
| | with others | 173 | 85.2 | 20(9.9%) | 153(75.4%) |
| Economic status | low | 24 | 11.8 | 2(1.0%) | 22(10.8%) |
| | medium | 152 | 74.9 | 16(7.9%) | 136(67.0%) |
| | rich | 27 | 13.3 | 4(2.0%) | 23(11.3%) |

Table 4.1 provides a summarized overview of sociodemographic variables for epilepsy patients at baseline, while also indicating the survival status (seizures-free or censored) for each category within these variables: There are a total of 203 epilepsy patients at baseline, with 119 (58.6%) were male and 84 (41.4%) were female. Out of the male patients, 118 (58.1%) were seizures-free, and 1 (0.5%) was censored. Among the female patients, 63 (31.0%) were seizures-free, and 21 (10.3%) were censored. The majority of patients are government employees (132 or 65.0%). Among farmers, 32 (15.8%) were seizures-free and 1 (0.5%) was censored. Among government employees, 114 (56.2%) were seizures-free and 18 (8.9%) were censored. The highest number of patients has a tertiary education level (127 or 62.6%). Among illiterate patients, 36 (17.7%) were seizures-free, and 3 (1.5%) were censored. Among those with a primary education, 14 (6.9%) were seizures-free and 1 (0.5%) was censored. Among those with a secondary education, 21 (10.3%) were seizures-free and 1 (0.5%) was censored. Among those with a tertiary education, 110 (54.2%) were seizures-free, and 17 (8.4%) were censored. There are more urban residents (140 or 69.0%) than rural residents (63 or 31.0%). Among rural residents, 56 (27.6%) were seizures-free and 7 (3.4%) were censored. Among urban residents, 125 (61.6%) were seizures-free and 15 (7.4%) were censored. The majority of patients are married (136 or 67.0%). Among divorced patients, 14 (6.9%) were seizures-free. Among married patients, 119 (58.6%) were seizures-free and 17 (8.4%) were censored. Among single patients, 33 (16.3%) were seizures-free and 4 (2.0%) were censored. Among widowed patients, 1 (0.5%) was seizures-free, and 14

(6.9%) were censored. Most patients live with others (173 or 85.2%). Among those living alone, 28 (13.8%) were seizures-free, and 2 (1.0%) were censored. Among those living with others, 153 (75.4%) were seizures-free, and 20 (9.9%) were censored. The economic status of patients varies, with the majority falling under the medium category (152 or 74.9%). Among low economic status patients, 22 (10.8%) were seizures-free and 2 (1.0%) were censored. Among medium economic status patients, 136 (67.0%) were seizures-free and 16 (7.9%) were censored. Among rich economic status patients, 23 (11.3%) were seizures-free, and 4 (2.0%) were censored. This interpretation provides insights into the distribution of epilepsy patients based on various sociodemographic variables and their survival status (seizures-free or censored) at baseline. It highlights potential patterns and associations between these variables and the patients' seizure outcomes.

Table 4.2: Summary of clinical variables

| Variables | Category | Frequency | Percentage | Status (Baseline) | |
|---|---|---|---|---|---|
| | | | | Censored | Seizures free |
| Type of seizure | generalized | 152 | 74.9 | 14(6.9%) | 138(68.0%) |
| | mixed | 28 | 13.8 | 1(0.5%) | 27(13.3%) |
| | partial | 23 | 11.3 | 7(3.4%) | 16(7.9%) |
| Family history of epilepsy | no | 123 | 60.6 | 2(1.0%) | 121(59.6%) |
| | yes | 80 | 39.4 | 20(9.9%) | 60(29.6%) |
| Sleep deprivation | no | 80 | 39.4 | 3(1.5%) | 77(37.9%) |
| | yes | 123 | 60.6 | 19(9.4%) | 104(51.2%) |
| Having stress | no | 82 | 40.4 | 22(10.8%) | 60(29.6%) |
| | yes | 121 | 59.6 | - | 121(59.6%) |
| Brain injury | no | 107 | 52.7 | 1(0.5%) | 106(52.2%) |
| | yes | 96 | 47.3 | 21(10.3%) | 75(36.9%) |
| Source of medication | free | 41 | 20.2 | - | 41(20.2%) |
| | payment | 162 | 79.8 | 22(10.8%) | 140(69.0%) |
| Epilepsy is controlled by modern drug | no | 71 | 35.0 | 5(2.5%) | 66(32.5%) |
| | yes | 132 | 65.0 | 17(8.4%) | 115(56.7%) |
| Type of AED | Carbamazepine | 95 | 46.8 | 18(8.9%) | 77(37.9%) |
| | Phenobarbitone | 33 | 16.3 | - | 33(16.3%) |

| | | | | | |
|---|---|---|---|---|---|
| | Phenytoin | 34 | 16.7 | - | 34(16.7%) |
| | Sodium Valporate | 25 | 12.3 | 3(1.5%) | 22(10.8%) |
| | more treatment | 16 | 7.9 | 1(0.5%) | 15(7.4%) |
| Alcohol consumption | no | 41 | 20.2 | - | 41(20.2%) |
| | yes | 162 | 79.8 | 22(10.8%) | 140(69.0%) |
| Exercise | no | 148 | 72.9 | 19(9.4%) | 129(63.5%) |
| | yes | 55 | 27.1 | 3(1.5%) | 52(25.6%) |
| Comorbidity | no | 120 | 59.1 | 3(1.5%) | 117(57.6%) |
| | yes | 83 | 40.9 | 19(9.4%) | 64(31.5%) |
| Chronic disease | combination | 109 | 53.7 | 12(5.9%) | 97(47.8%) |
| | diabetic melitus | 9 | 4.4 | 1(0.5%) | 8(3.9%) |
| | HIV | 12 | 5.9 | - | 12(5.9%) |
| | hypertension | 16 | 7.9 | 1(0.5%) | 15(7.4%) |
| | no | 38 | 18.7 | 6(3.0%) | 32(15.8%) |
| | stroke | 19 | 9.4 | 2(1.0%) | 17(8.4%) |
| Adherence level | low | 35 | 17.2 | 3(1.5%) | 32(15.8%) |
| | medium | 53 | 26.1 | 6(3.0%) | 47(23.2%) |
| | high | 115 | 56.7 | 13(6.4%) | 102(50.2%) |

Most patients have generalized seizures (152 or 74.9%). Among those with generalized seizures, 138 (68.0%) were seizures-free, and 14 (6.9%) were censored. There are patients with mixed seizures (28 or 13.8%), with 27 (13.3%) seizures-free and 1 (0.5%) censored. Patients with partial seizures are 23 (11.3%), out of which 16 (7.9%) were seizures-free, and 7 (3.4%) were censored. Most patients have no family history of epilepsy (123 or 60.6%). Among patients with no family history, 121 (59.6%) were seizures-free, and 2 (1.0%) were censored. Among patients with a family history, 60 (29.6%) were seizures-free, and 20 (9.9%) were censored. More patients have sleep than those without it (123 or 60.6% vs. 80 or 39.4%). Among patients with sleep, 104 (51.2%) were seizures-free, and 19 (9.4%) were censored. Among patients without sleep, 77 (37.9%) were seizures-free, and 3 (1.5%) were censored. More patients reported stress (121 or 59.6%) compared to those without it (82 or 40.4%). All patients without stress were seizures-free (60 or 29.6%). Among patients with stress, the status is not specified, and all of them (121 or 59.6%) are listed under the seizures-free category.

More patients do not have a reported brain injury (107 or 52.7%). Among patients without a brain injury, 106 (52.2%) were seizures-free, and 1 (0.5%) was censored. Among patients with a brain injury, 75 (36.9%) were seizures-free, and 21 (10.3%) were censored. The majority of patients receive payment-based treatment 162(79.8%). Among patients with payment-based treatment, 140 (69.0%) were seizures-free, and 22 (10.8%) were censored. More patients are controlled by modern drugs (132 or 65.0%). Among patients not controlled by modern drugs, 115 (56.7%) were seizures-free, and 17 (8.4%) were censored. Among patients controlled by modern drugs, 66 (32.5%) were seizures-free, and 5 (2.5%) were censored. The most common AED is Carbamazepine (95 or 46.8%). Other types of AEDs are also present but without specified survival status. More patients use alcohol (162 or 79.8%) compared to those who do not (41 or 20.2%). Among alcohol users, 140 (69.0%) were seizures-free, and 22 (10.8%) were censored. The majority of patients do not exercise (148 or 72.9%). Among those who exercise, 52 (25.6%) were seizures-free, and 3 (1.5%) were censored. More patients do not have reported co-morbidities (120 or 59.1%). Among patients without co-morbidities, 117 (57.6%) were seizures-free, and 3 (1.5%) were censored. Among patients with co-morbidities, 64 (31.5%) were seizures-free, and 19 (9.4%) were censored. The most common chronic disease is a combination of conditions (109 or 53.7%). Other specific chronic diseases are present but with varying prevalence. Most patients' response high adherence level (115 or 56.7%). Among those with low adherence 35(17.2%), 32(15.8%) were seizures-free, and 3(1.5%) were censored. There are patients' response medium adherence (53or 26.1%), with 47(23.2%) seizures-free and 6(3.0%) censored. Patients' response high adherence level is (115or 56.7%)., out of which 102(50.2%) were seizures-free, and 13(6.4%) were censored.

Table 4.3: Mean and standard deviation of seizure attack measurement with event (seizure free) and censoring status of epileptic patients at HUCSH

| Status | Measurement | Baseline | Visit1 | Visit2 | Visit3 | Visit4 | Visit5 |
|---|---|---|---|---|---|---|---|
| | | 0 | 12 | 24 | 36 | 48 | 60 |
| Seizure free | Mean | .73 | .08 | .00 | .00 | .01 | .00 |
| | Std. Deviation | 2.354 | .500 | .000 | .000 | .084 | .000 |
| Censored | Mean | 5.31 | 3.20 | 2.14 | 1.33 | 1.22 | .18 |
| | Std. Deviation | 2.628 | 1.765 | 1.060 | .555 | .485 | .528 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Total | Mean | 4.82 | 2.65 | .96 | .18 | .25 | .04 |
| | Std. Deviation | 2.962 | 2.008 | 1.281 | .496 | .541 | .267 |

Table 4.3 represents the summary of longitudinal seizure episode counts and seizure freedom survival outcomes over multiple visits. The data is presented in terms of mean (average) and standard deviation (a measure of the dispersion or spread of data) for different measurements (Seizure free, Censored, and Total) at different time points (Baseline, Visit 1, Visit 2, Visit 3, Visit 4, Visit 5).

The "Seizure Free" category refers to individuals who are free from seizures. At the baseline visit (time point 0), the mean number of seizure-free individuals is 0.73. This means that, on average, less than 1 person is seizure-free at the beginning. As time progresses, the number of seizure-free individuals significantly drops, reaching close to 0 at Visit 2 and remaining near 0 for the subsequent visits. The standard deviation indicates the variability in the number of seizure-free individuals. It decreases over visits, suggesting a more consistent lack of seizure-free cases.

The "Censored" category refers to individuals whose data might be incomplete or not fully captured. This can occur due to various reasons, such as lost to follow-up or withdrawal from the study. At the beginning of the study, participants experienced an average of 5.31 censored episodes and the mean number of censored cases decreases as time goes on. This could imply that over time, fewer individuals have incomplete data. The standard deviation shows the variability in the number of censored cases. It decreases, indicating that the variability is reducing over time.

The mean total count decreases over time, suggesting that the number of individuals in the study decreases as time progresses. The standard deviation in the total count also decreases over time, indicating less variability in the number of individuals participating in the study.

In summary, the data suggests that the number of seizure-free individuals decreases significantly over time, reaching almost zero by the second visit. The number of censored cases also decreases over time. Overall, the total number of individuals in the study decreases over time. This interpretation gives an overview of the changes in seizure episodes, seizure freedom, and participant count over the visits in the study.

### 4.2 Separate analysis for survival model

Non-parametric analysis for survival data of epileptic patients

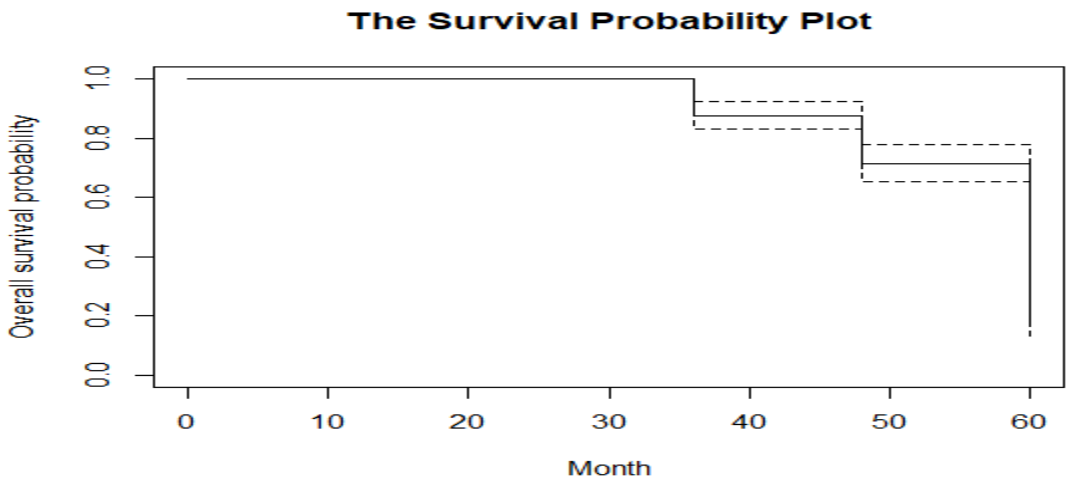



Figure 4.1: Plot of Survival Probability of Patients

The Kaplan-Meier survival plot shows the estimated survival probabilities over time for the group of individuals with seizure data. The x-axis represents time in months, and the y-axis represents the cumulative survival probability.

The curve starts at 1, indicating that initially, all individuals are seizure-free. As time progresses, the curve gradually declines, representing the decreasing probability of remaining seizure-free. The steps in the curve correspond to the occurrence of seizures (events). Plateaus in the curve indicate periods when no seizures occurred, and the survival probability remained constant. By observing the shape of the curve and the timing of the steps, we can gain insights into the overall seizure-free survival experience of the group. Steeper declines indicate higher event rates (seizures), while longer plateaus suggest periods of relative stability without seizures.

**Kaplan- Meier survival curves for categorical variables of epileptic patients**

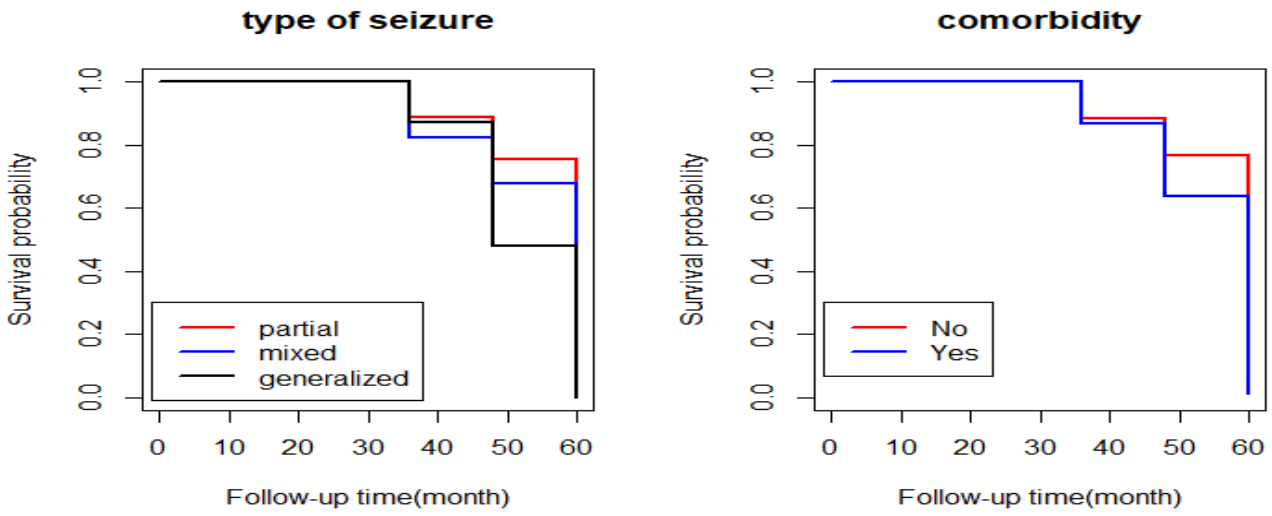



***Figure 4.2 Kaplan- Meier survival curves for types of seizure and comorbidity***

The survival time of partial seizure type patients had higher survival time compared to mixed and generalized seizure type. Being generalized seizure type patient experience the event seizure free within a short period of time compared to mixed and partial epileptic patients.

According to the Kaplan-Meier survival plot for comorbidity, epileptic patients with no comorbidities had a higher chance of seizure freedom, indicating a shorter time duration to achieve seizure-free status.

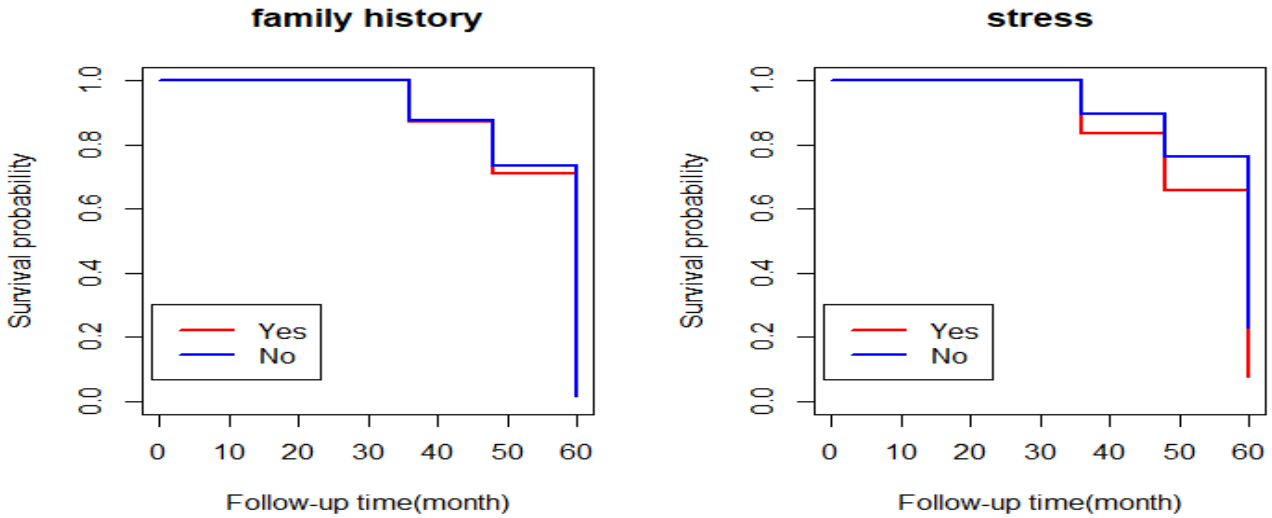


*Figure 4.3 Kaplan- Meier survival curves for family history and stress*

The Kaplan-Meier curve for family history indicates that patients without a family history of seizures experienced longer survival times compared to those with a family history. Consequently, patients with a family history of seizures had poorer survival outcomes, with a lower likelihood of achieving seizure freedom within the same time frame. Similarly, patients without stress experienced higher survival rates compared to those with stress.

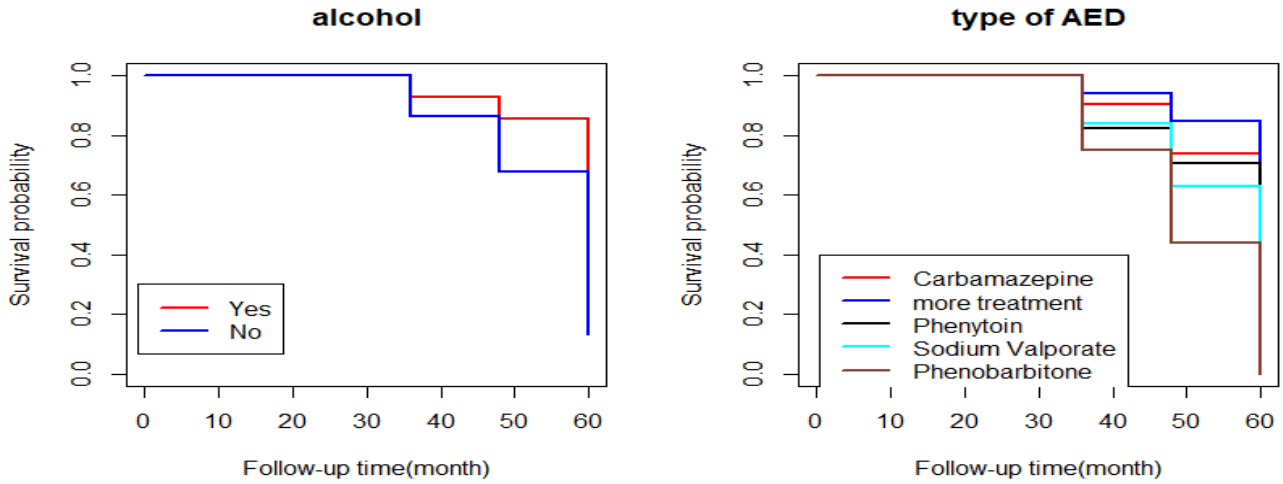



Figure 4.4 Kaplan- Meier survival curves for alcohol and types of AED

From the Kaplan Meier survival plot of alcohol, epileptic patients who are no alcohol had higher survival chance of seizure freedom, means that they took short time duration to experience the event seizure free.

From the Kaplan Meier survival plot of types of AED, epileptic patients who take more treatment had higher survival chance of seizure freedom than other category, means that they took short time duration to experience the event seizure free.

**Log-rank test for each categorical variable**

Table 4.4 Comparison of survival experience of epilepsy patients using several explanatory variables

| **Variables** | **Log-Rank** | | |
|---|---|---|---|
| | **Chi-Square** | **DF** | **Pr>Chi-Square** |

| | | | |
|---|---|---|---|
| Type of AED | 28 | 4 | 1e-05 |
| Type of seizure | 9.6 | 2 | 0.008 |
| Comorbidity | 18.2 | 1 | 2e-05 |
| Family History | 11.4 | 1 | 7e-04 |
| Stress | 2.9 | 1 | 0.09 |
| Alcohol | 9.2 | 1 | 0.002 |

Table 4.4 represents the results of the log-rank test, which is a statistical test used to compare survival curves between different categories of a variable. The test assesses whether the observed differences in survival between these categories are statistically significant.

The p-value for the log-rank test comparing different treatment types (Phenobarbitone, Phenytoin, Sodium Valporate, more treatment) is very small (1e-05), indicating a highly significant difference in survival curves between these treatment types. This suggests that the choice of antiepileptic drug (AED) treatment significantly affects the survival experience of individuals with respect to seizures. The p-value for the log-rank test comparing seizure types (mixed vs. partial) is 0.008, indicating a statistically significant difference in survival curves between these two types of seizures. This suggests that the type of seizure (mixed or partial) has an impact on the survival experience of individuals. The p-value for the log-rank test comparing individuals with and without comorbidity is very small (2e-05), indicating a highly significant difference in survival curves between these two groups. This suggests that the presence of comorbidity significantly influences the survival experience of individuals. The p-value for the log-rank test comparing individuals with and without a family history of seizures is 0.0007, indicating a statistically significant difference in survival curves between these groups. This suggests that having a family history of seizures impacts the survival experience of individuals. The p-value for the log-rank test comparing individuals experiencing stress and those not experiencing stress is 0.09, which is not statistically significant (commonly considered significant if less than 0.05). This suggests that stress may not have a significant impact on the survival experience of individuals. The p-value for the log-rank test comparing individuals who consume alcohol and those who don't is 0.002, indicating a statistically significant difference in

survival curves between these two groups. This suggests that alcohol consumption has an impact on the survival experience of individuals.

**Checking Assumptions for Proportional Hazard**

Table 4.6 is useful to test whether the hazard of the hypertension disease is proportional (that means its risk is not timely varied) or not. And we test the hypothesis hazard is proportional versus not proportional. The significance of the test (all statistical tests used in this study) is valid at 5% level of significance.

Table 4.6: Test of Proportional Hazards

| Variable | Chi-square ($\chi^2$) | df | p-value |
|---|---|---|---|
| Type of AED | 0.5214 | 4 | 0.97 |
| Type of seizure | 1.3593 | 2 | 0.51 |
| comorbidity | 0.5407 | 1 | 0.46 |
| Family history | 1.7549 | 1 | 0.19 |
| stress | 0.0151 | 1 | 0.90 |
| alcohol | 0.2945 | 1 | 0.59 |
| GLOBAL | 8.9707 | 11 | 0.62 |

Proportional hazard at each variable since p-value at each covariate is larger than 0.05. Thus, the proportional hazard assumption is not violated.

**Model Comparison**

Before moving on to other analyses, we had compared the semi-parametric Cox PH, and parametric models (exponential, Weibull, log-normal and log-logistic models) using their information criteria.

Table 4.7: Model comparison for Bayesian survival model

| No | Model | DIC | WAIC | Marginal log-Likelihood |
|---|---|---|---|---|
| 1 | Cox-proportional model | 1596.20 | 4974.73 | -350.32 |

| | | | | |
|---|---|---|---|---|
| 2 | Exponential | 1697.63 | 1689.15 | -888.81 |
| 3 | Weibull | 1205.74 | 1202.38 | -649.88 |
| 4 | Lognormal | 1277.63 | 1278.69 | -704.30 |

The models being compared are semi-parametric Cox PH, Exponential, Weibull, and Lognormal survival models, and the comparison is based on three criteria: Deviance Information Criterion (DIC), Widely Applicable Information Criterion (WAIC), and Marginal log-Likelihood.

The Weibull Model has the lowest values for both DIC and WAIC. Hence it was the best model and we had used it to analyze the survival data Table 4.7.

Table 4.8 Bayesian Multivariable analysis for Weibull parametric model

| Variable | | Post. Mean | St.dev | 95% Cred.I |
|---|---|---|---|---|
| Intercept | | -35.860 | 0.391 | (-36.627, -35.094) |
| Type of AED | Phenobarbitone; *ref* Carbamazepine | -0.139 | 0.314 | (-0.755, 0.478) |
| | Phenytoin | 0.946 | 0.276 | (0.404, 1.488) |
| | Sodium Valporate | 0.335 | 0.261 | (-0.177, 0.847) |
| | more treatment | 1.337 | 0.353 | (0.644, 2.029) |
| Type of seizure | Mixed; *ref* generalized | 0.192 | 0.253 | (-0.304, 0.689) |
| | partial | 0.309 | 0.241 | (-0.164, 0.783) |
| comorbidity | Yes ; *ref* No | -0.840 | 0.208 | (-1.248, -0.432) |
| Family history | Yes ; *ref* No | -0.262 | 0.218 | (-0.689, 0.165) |
| Stress | Yes ; *ref* No | -0.010 | 0.212 | (-0.425, 0.405) |

| | | | | |
|---|---|---|---|---|
| Alcohol | Yes ; *ref* No | 0.170 | 0.305 | (-0.428, 0.768) |

Table 4.8 appears to be a model of Weibull survival using the INLA (Integrated Nested Laplace Approximations) package. The table displays the estimates of the model's coefficients along with their standard deviations and 95% credible interval.

Phenobarbitone, with a mean coefficient of -0.139, exhibits a negative effect on the hazard, but the 95% credible interval (-0.755, 0.478) includes zero, indicating potential non-significance. Phenytoin, with a mean coefficient of 0.946, shows a statistically significant positive effect on the hazard (95% credible interval: 0.404, 1.488). Sodium Valproate, while having a mean coefficient of 0.335, does not demonstrate statistical significance (95% credible interval: -0.177, 0.847). Patients receiving more than one treatment exhibit a positive effect on the hazard, with a mean coefficient of 1.337. The 95% credible interval (0.644, 2.029) excludes zero, suggesting statistical significance.

The type of seizure variable represents the type of seizure experienced by the patient. The specific seizure types included in the model are "mixed" and "partial." Each seizure type has an associated coefficient indicating its effect on the hazard rate relative to the baseline hazard. However, the 95% credible intervals for both "mixed" (-0.304, 0.689) and "partial" (-0.164, 0.783) include zero, suggesting that the effects of these seizure types may not be statistically significant.

The Co-morbidity variable has one category representing whether a patient has co-morbidities or not. The mean is -0.840, and the credible interval (-1.248, -0.432) does not include zero, suggesting that having co-morbidities has a statistically significant negative effect on seizure freedom.

The Family history variable has one category representing whether a patient has Family history of epilepsy or not. The mean is -0.262, but the credible interval (-0.689, 0.165) includes zero, indicating that the effect of febrile convulsions may not be statistically significant.

The Stress variable has one category representing whether a patient experiences stress or not. The mean is -0.010, and the credible interval (-0.425, 0.405) includes zero, suggesting that the effect of stress is statistically inconclusive.

The Alcohol variable has one category representing whether a patient consumes alcohol or not. The mean is 0.170, and the credible interval (-0.428, 0.768) includes zero, indicating that the effect of alcohol consumption may not be statistically significant.

**Model diagnostic tests for survival model**

In the INLA method density plots is one of the methods of assessing the diagnosis of the estimation. Therefore, the density plots in (Appendix B, Figure B.1.) were nearly similar to the normal plot. This result indicated that the survival model fitted the data well and it was converged.

### 4.3 Separate Analysis of the Longitudinal Data

**Exploring the Longitudinal Data**

Exploratory data analysis was conducted in order to investigate various associations, structures and patterns exhibited in the data set. Additionally, the individual profile plots mean structure plots, and variance plots were obtained in order to gain some insights of the data (Verbeke et al., 2010).

Individual Profile Plots of Epileptic Patients With longitudinal or repeated measures data, there are often two aspects that are interesting. First, how much variability is there between individual units at a given time or measure? Second, how much variance is their within units over time or measure? One convenient way to visualize both is using a spaghetti plot. This draws a plot with separate lines for each unit. The space between lines represents between unit variability, the change in each line (slope) represents within variability. Plotting observed profiles over time helps identify general trends within subjects and may detect change over time that provides information about the variability at given times.

Connecting the repeated measurements for each epileptic seizure patients over time shows there is a discernible pattern common to most patients. These individual profiles can also provide some information on between epileptic patients variability and illustrate that there is change among patients over time and the variability of seizure attacks for 12 month seems larger at the beginning compared to the end and over all means of variance decreased over time shown in Figure 4.2 at the right side. Individual profiles plot tends to show that there is variability between

Average seizure attacks measurements and variability within patients. That is, Average seizure attack levels for some patients were going down, and others were going up over the time points. This shows that, not all the patients responded the same to Anti epileptic drug (AEDs) treatment, and thus suggests the relevance of a mixed effects model to address the random effects part in addition to fixed effects.

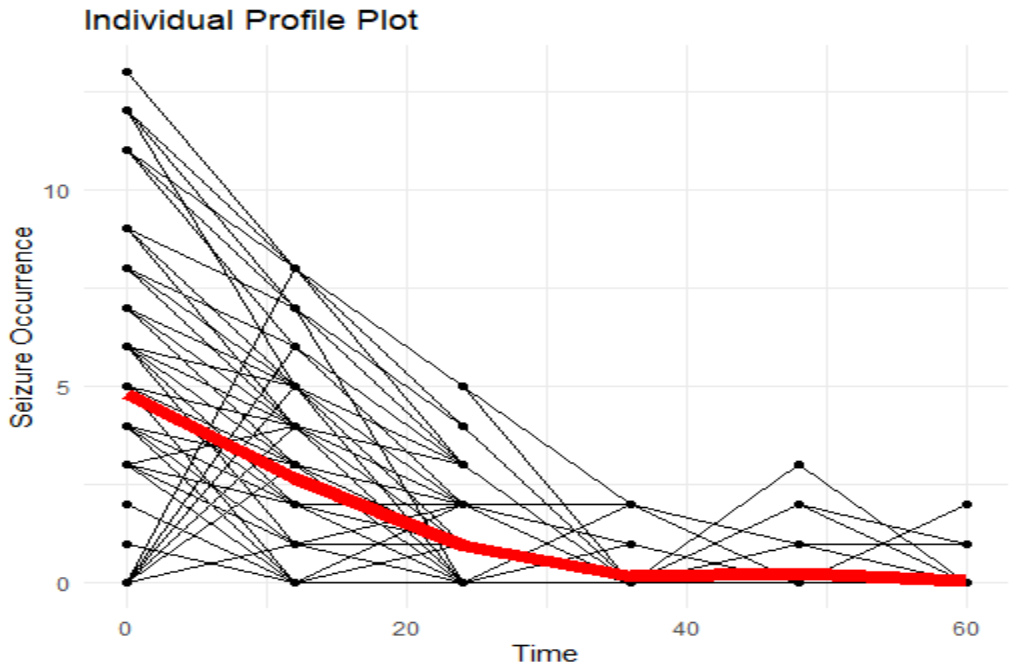



***Figure 4.5 Individual and Mean profile plot for the number of seizure attacks***

The red line Loess smoothing plot technique shows the mean structure of seizure attacks measurements decrease over time with a linear relationship (i.e., the relationship between seizure attacks and measurement time linear) and finally seems constant.

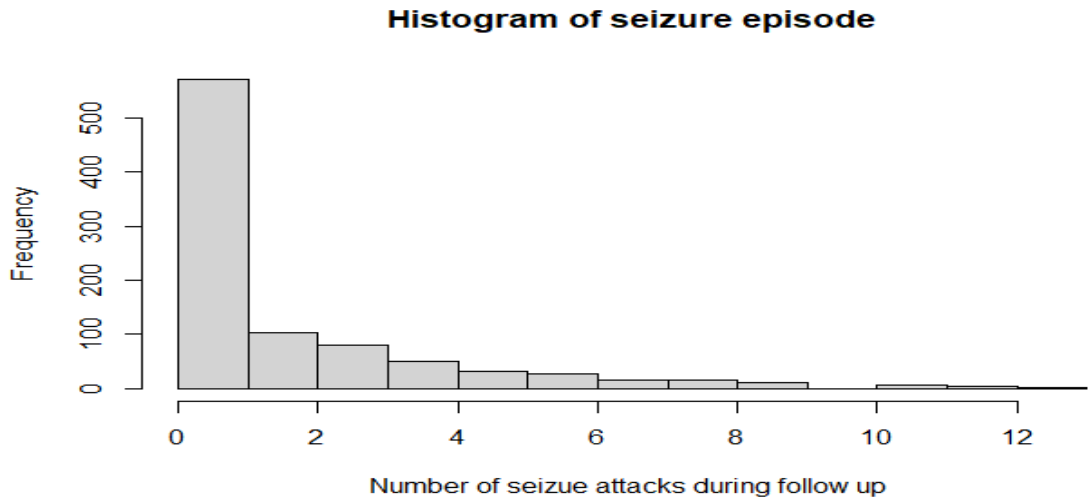



***Figure 4.6 Frequency distribution of the number of seizure attacks***

The above histogram plot that shows the frequency distribution of the values in the count longitudinal data. Since the data seems to contain a mix of non-zero values and a substantial number of zero values, the histogram may exhibit a skewed distribution with a long tail towards zero. This could indicate the presence of excess zeros or zero inflation in the data. A significant peak at zero could suggest zero inflation, which means that there is an excess of zero values beyond what would be expected from a standard count distribution.

**Generalized Linear Mixed Effect Model**

Table 4.9 Model comparison for count longitudinal outcome

| No | Model | DIC | WAIC | Marginal log-Likelihood |
|---|---|---|---|---|
| 1 | Poisson | 4805.77 | 5030.12 | -2509.73 |
| 2 | Negative binomial | 3725.17 | 3720.77 | -1942.55 |
| 3 | Poisson zero inflated | 3634.51 | 3665.78 | -1911.12 |
| 4 | Negative binomial zero inflated | 3601.65 | 3601.86 | -1881.75 |

The Negative Binomial Zero-Inflated Model has the lowest values for DIC, WAIC, and Marginal log-Likelihood. Therefore, it seems to be the best-fitting model among the options provided. The Negative Binomial Model has the second-best fit, followed by the Poisson Zero-Inflated Model, and finally the Poisson Model, which has the highest values for all three criteria.

Table 4.10 Bayesian multivariable analysis for Zero Inflated Negative Binomial model

| Variables | | Post. Mean | St.dev | 95% Cred.I |
|---|---|---|---|---|
| Intercept | | 0.394 | 0.189 | (0.024, 0.764) |
| Type of AED | Phenobarbitone; *ref* Carbamazepine | -0.093 | 0.136 | (-0.360, 0.174) |
| | Phenytoin | 0.451 | 0.108 | (0.240, 0.663) |
| | Sodium Valporate | -0.002 | 0.113 | (-0.224, 0.219) |
| | more treatment | 0.328 | 0.143 | (0.049, 0.608) |
| Type of seizure | mixed; *ref* generalized | 0.070 | 0.096 | (-0.117, 0.258) |
| | partial | 0.332 | 0.101 | (0.133, 0.531) |
| Exercise | Yes; *ref* No | 0.519 | 0.122 | (0.280, 0.758) |
| Comorbidity | Yes ; *ref* No | 0.340 | 0.086 | (0.173, 0.508) |
| Family history of epilepsy | Yes; *ref* No | 0.435 | 0.083 | (0.273, 0.598) |
| Brain injury | Yes; *ref* No | 0.019 | 0.095 | (-0.167, 0.204) |
| Alcohol consumption | Yes; *ref* No | 0.678 | 0.161 | (0.363, 0.993) |
| Chronic disease | Diabetic melitus; *ref* | -0.055 | 0.156 | (-0.361, 0.250) |

| | | | | |
|---|---|---|---|---|
| | combination | | | |
| | HIV | 0.130 | 0.130 | (-0.124, 0.385) |
| | hypertension | -0.105 | 0.128 | (-0.357, 0.146) |
| | no | -0.205 | 0.090 | (-0.382, -0.028) |
| | stroke | -0.217 | 0.120 | (-0.451, 0.018) |
| Zero inflated | size for nbinomial zero-inflated observations | 4.79 | 1.02 | (3.160,7.17) |
| | zero-probability parameter | 5.54e-01 | 1.50e-02 | (0.525,5.84e-01) |

In table 4.11, the results of a zero-inflated negative binomial analysis for factors affecting the fre quency of seizure attacks in epilepsy patients are presented. The analysis includes various variabl es, and each row represents a different category or level within that variable. The table provides i nformation about the mean, standard deviation (St.dev), and 95% credible intervals (CI) for each category's effect on the frequency of seizure attacks. In Bayesian analysis, the credible interval is used to express uncertainty about the parameter estimates, and it's analogous to the frequentist co nfidence interval.

Phenobarbitone, with a mean effect of -0.093, exhibits a non-significant impact on seizure freque ncy, as the 95% credible interval (-0.360, 0.174) includes zero. Phenytoin, on the other hand, de monstrates a statistically significant positive effect on increasing the frequency of seizure attacks , with a mean effect of 0.451 and a credible interval (0.240, 0.663) that excludes zero. Sodium V alproate presents an inconclusive effect on seizure frequency, as its mean effect is -0.002, and the 95% credible interval (-0.224, 0.219) includes zero. Patients receiving more treatments show a st atistically significant positive effect on increasing seizure frequency, with a mean effect of 0.328 and a credible interval (0.049, 0.608) that excludes zero.

In terms of seizure types, a mixed seizure type exhibits a non-significant effect on seizure freque ncy, with a mean effect of 0.070 and a credible interval (-0.117, 0.258) including zero.

In contrast, patients with a partial seizure type have a statistically significant higher frequency of seizure attacks compared to the baseline, as indicated by a mean effect of 0.332 and a credible int erval (0.133, 0.531) that excludes zero.

The exercise variable has one category representing whether a patient exercises or not. The mean is 0.519, and the credible interval (0.280, 0.758) does not include zero, suggesting that patients w ho exercise have a statistically significant lower frequency of seizure attacks.

The Co-morbidity variable has one category representing whether a patient has co-morbidities or not. The mean is 0.340, and the credible interval (0.173, 0.508) does not include zero, indicating that having co-morbidities has a statistically significant positive effect on the frequency of seizur e attacks.

The Family history variable has one category representing whether a patient has a family history of epilepsy or not. The mean is 0.435, and the credible interval (0.273, 0.598) does not include ze ro, suggesting that having a family history of epilepsy has a statistically significant positive effec t on the frequency of seizure attacks.

The variable Brain injury has one category representing whether a patient has a history of brain i njury or not. The mean is 0.019, and the credible interval (-0.167, 0.204) includes zero, making t he effect of brain injury statistically inconclusive.
The variable alcohol has one category representing whether a patient consumes alcohol or not. T he mean is 0.678, and the credible interval (0.363, 0.993) does not include zero, indicating alcoh ol consumption has a statistically significant positive effect on the frequency of seizure attack.

Diabetic Melitus exhibits an inconclusive effect on seizure frequency, with a mean effect of -0.055 and a credible interval (-0.361, 0.250) that includes zero. HIV presents a non-significant e ffect on seizure frequency, as indicated by a mean effect of 0.130 and a credible interval (-0.124, 0.385) that includes zero. Hypertension demonstrates an inconclusive impact on seizure frequenc y, with a mean effect of -0.105 and a credible interval (-0.357, 0.146) that includes zero. Patients without any chronic disease have a statistically significant lower frequency of seizure attacks co

mpared to the baseline, as evidenced by a mean effect of -0.205 and a credible interval (-0.382, -0.028) that excludes zero.

Zero-Inflated Size for Negative Binomial Zero-Inflated Observations: The estimated posterior mean is 4.79, with a standard deviation of 1.02. The 95% credible interval (CI) for this parameter is (3.160, 7.17). In the context of a zero-inflated negative binomial model, this parameter is often associated with the dispersion of the negative binomial distribution for non-zero counts.

Zero-Probability Parameter: The estimated posterior mean for the zero-probability parameter is 0.554, with a standard deviation of 0.015. The 95% credible interval for this parameter is (0.525, 0.584). This parameter represents the probability of observing a zero count, and a higher value in dicates a higher probability of excess zeros in the data.

**Diagnostic for posterior mean random effects**

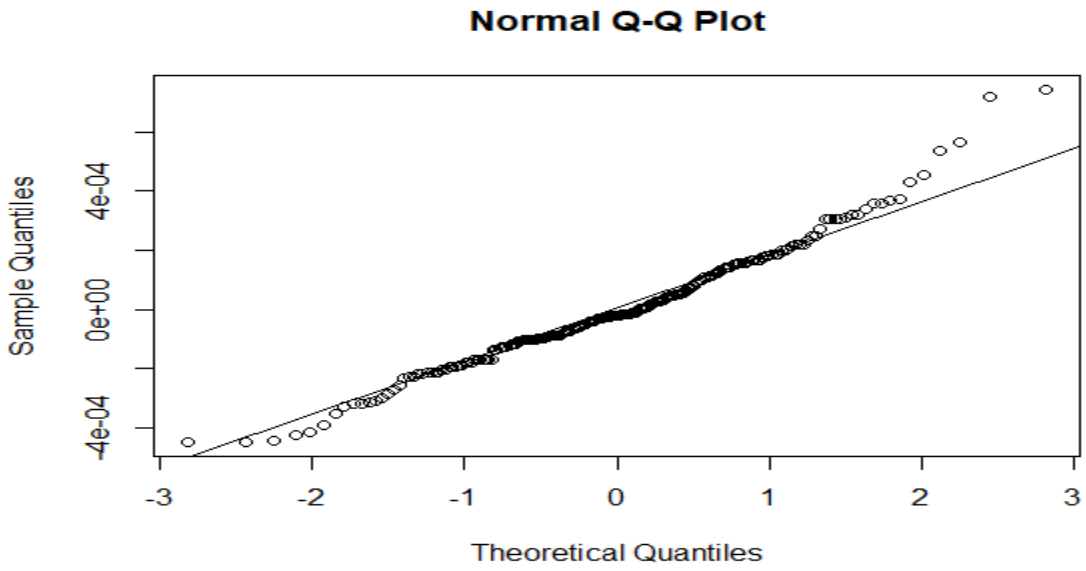



***Figure 4.7 QQ plot of the posterior mean random effects***

As we expect these posterior means to be approximately normal, a QQ plot is a natural graphical check so the posterior mean random effects approximately normal.

**Model diagnostic tests for longitudinal model**

In (Appendix B, Figure B.3.) the density plot for zero-inflated negative binomial were presented. The density plot was approximately similar to the normal plot. This result indicated that the longi tudinal model fitted the data well and it was converged.

### 4.4 Bayesian Joint Longitudinal-Survival Model Analysis

The inla() function's major arguments were the previously estimated the longitudinal and time-to -event models. All estimated parameters' posterior mean were reported, along with their 95% cre dible interval (Table 4.11).

Table 4.11: Summary of Bayesian joint model

| Zero-inflated negative binomial | | | | |
|---|---|---|---|---|
| Variable | Category | Post. Mean | St.dev | 95% Cred.I |
| Intercept | - | -0.833 | 0.654 | (-2.135, 0.452) |
| Type of AED | Phenobarbitone; *ref* Carbamazepine | -0.054 | 0.122 | (-0.294, 0.186) |
| | Phenytoin | 0.411 | 0.096 | (0.222, 0.6000) |
| | Sodium Valporate | -0.036 | 0.096 | (-0.225, 0.152) |
| | more treatment | 0.274 | 0.126 | (0.027, 0.521) |
| Type of seizure | mixed; *ref* generali zed | -0.026 | 0.084 | (-0.191, 0.140) |
| | partial | 0.265 | 0.083 | (0.101, 0.429) |
| comorbidity | Yes ; *ref* No | 0.344 | 0.074 | (0.198, 0.489) |
| Exercise | Yes ; *ref* No | 0.529 | 0.103 | (0.328, 0.731) |
| Family history | Yes; *ref* No | 0.417 | 0.071 | (0.277, 0.557) |
| Brain injury | Yes ; *ref* No | 0.084 | 0.083 | (-0.079, 0.247) |
| Alcohol | Yes ; *ref* No | 0.653 | 0.140 | (0.378, 0.927) |
| Chronic disease | diabetic melitus; *ref* combination | -0.043 | 0.133 | (-0.303, 0.218) |
| | HIV | 0.106 | 0.111 | (-0.112, 0.324) |
| | hypertension | -0.098 | 0.112 | (-0.317, 0.122) |
| | no | -0.186 | 0.078 | (-0.339, -0.034) |
| | stroke | -0.204 | 0.103 | (-0.405, -0.003) |

| | | | | |
|---|---|---|---|---|
| size for nbinomial zero-inflated observations | - | 1.37e+02 | 7.24e+01 | (29.736, 3.01e+02) |
| zero-probability parameter | - | 5.57e-01 | 1.70e-02 | (0.526, 5.95e-01) |
| **Weibull model** | | | | |
| Intercept | | -15.181 | 0.455 | (-16.018,14.233) |
| Type of AED | Phenobarbitone; *ref* Carbamazepine | -0.054 | 0.122 | (-0.294, 0.186) |
| | Phenytoin | 0.411 | 0.096 | (0.222, 0.6000) |
| | Sodium Valporate | -0.036 | 0.096 | (-0.225, 0.152) |
| | more treatment | 0.274 | 0.126 | (0.027, 0.521) |
| Type of seizure | mixed; *ref* generalized | 0.093 | 0.252 | (-0.401, 0.587) |
| | partial | 0.125 | 0.244 | (-0.354, 0.604) |
| Comorbidity | Yes ; *ref* No | -0.685 | 0.199 | (-1.076, 1.114) |
| Family history | Yes ; *ref* No | -0.324 | 0.218 | (-0.771, -0.294) |
| Stress | Yes ; *ref* No | 0.005 | 0.205 | (-0.397, 0.407) |
| Alcohol | Yes ; *ref* No | 0.080 | 0.299 | -0.507, 0.667) |
| Association parameter | - | 0.604 | 0.205 | (0.121, 0.907) |

Table 4.11 represents a joint model for zero-inflated negative binomial (ZINB) and Weibull survival models of seizure attack and seizure freedom for epilepsy patients. The model has been fitted using the INLA (Integrated Nested Laplace Approximations) package. The table displays the estimates of the model's coefficients along with their standard deviations and 95% credible intervals.

**Zero-Inflated Negative Binomial Model Part**

Phenobarbitone is associated with a potential reduction in seizure attack counts, as indicated by a mean coefficient of -0.054. However, the 95% credible interval (-0.294, 0.186) includes zero, suggesting that the effect may not be statistically significant. Phenytoin, with a mean coefficient of 0.411, is significantly associated with an increase in seizure attack counts, supported by a 95% credible interval of (0.222, 0.600). Sodium Valproate, with a mean coefficient of -0.036, does not show a statistically significant association with seizure attack counts, as the 95% credible

interval (-0.225, 0.152) includes zero. Receiving more than one treatment is associated with a potential increase in seizure attack counts, with a mean coefficient of 0.274. The 95% credible interval (0.027, 0.521) excludes zero, indicating statistical significance.

For seizure types, having a mixed seizure type is not significantly associated with the count of seizure attacks, as suggested by a mean coefficient of -0.026 and a 95% credible interval of (-0.191, 0.140) that includes zero. In contrast, having a partial seizure type is associated with a potential increase in seizure attack counts, with a mean coefficient of 0.265. The 95% credible interval (0.101, 0.429) supports this association.
The variable Co-morbidity has one category representing whether a patient has co-morbidities or not. The mean is 0.344, and the credible interval (0.198, 0.489) does not include zero, indicating that having co-morbidities has a statistically significant positive effect on the frequency of seizur e attacks.

The variable Family history has one category representing whether a patient has a family history of epilepsy or not. The mean is 0.417, and the credible interval (0.277, 0.557) does not include ze ro, suggesting that having a family history of epilepsy has a statistically significant positive effec t on the frequency of seizure attacks.

The variable Exercise has one category representing whether a patient do exercise or not. The me an is 0.529, and the credible interval (0.328, 0.731) does not include zero, suggesting that doing e xercise has a statistically significant positive effect on the frequency of seizure attacks.

The variable Alcohol has one category representing whether a patient have a history of alcohol c onsumption or not. The mean is 0.653, and the credible interval (0.378, 0.927) does not include z ero, suggesting that having a history of alcohol consumption has a statistically significant positiv e effect on the frequency of seizure attacks.

Diabetic melitus is not significantly associated with the longitudinal outcome, as indicated by a mean coefficient of -0.043. The 95% credible interval (-0.303, 0.218) includes zero, suggesting a lack of statistical significance. Similarly, having HIV is not significantly associated with the lon gitudinal outcome, with a mean coefficient of 0.106 and a 95% credible interval of (-0.112, 0.324 ) that includes zero. The presence of hypertension is not significantly associated with the longitu dinal outcome, as suggested by a mean coefficient of -0.098 and a 95% credible interval of (-0.31

7, 0.122) that includes zero. Patients with no chronic disease exhibit a potential decrease in the lo ngitudinal outcome, with a mean coefficient of -0.186. The 95% credible interval (-0.339, -0.034 ) excludes zero, indicating statistical significance. Having a stroke is associated with a potential d ecrease in the longitudinal outcome, as suggested by a mean coefficient of -0.204. The 95% credi ble interval (-0.405, -0.003) excludes zero, supporting the statistical significance of this associati on.

Brain injury variables represent the presence or absence of certain conditions or factors. Each of t hese variables has a coefficient indicating its effect on the count data, but none of them appear to be statistically significant as their 95% credible intervals include zero.

Zero-Probability Parameter: The mean of the zero-probability parameter is 0.557, which represents the estimated probability of observing zero counts (no seizures) when all predictor variables are at their reference levels.

**Weibull Survival Model Part**

Phenobarbitone is not significantly associated with the survival outcome of seizure freedom, as indicated by a mean coefficient of -0.185. The 95% credible interval (-0.805, 0.436) includes zero, suggesting a lack of statistical significance. In contrast, Phenytoin is associated with a statistically significant increase in the hazard of achieving seizure freedom, with a mean coefficient of 0.828 and a 95% credible interval of (0.309, 1.346). Sodium Valproate does not show a statistically significant association with the survival outcome, as suggested by a mean coefficient of 0.343 and a 95% credible interval of (-0.172, 0.859) that includes zero. Patients receiving more than one treatment exhibit a potential increase in the hazard of achieving seizure freedom, with a mean coefficient of 1.111. The 95% credible interval (0.426, 1.797) excludes zero, indicating statistical significance.

Regarding seizure types, having a mixed or partial type of seizure is not significantly associated with the survival outcome, as indicated by mean coefficients of 0.093 and 0.125, respectively, and 95% credible intervals that include zero.

For Comorbidity variable the mean coefficient is -0.685, with a 95% credible interval of (-1.076, -0.294). It suggests that having a comorbidity is significantly associated with the survival outcome (seizure freedom).

Family history, Stress, Alcohol variables represent the presence or absence of certain conditions or factors, similar to the ZINB model. Each of these variables has a coefficient indicating its effect on the survival outcome, but none of them appear to be statistically significant as their 95% credible intervals include zero.

Association Parameter: The mean of the association parameter is 0.604, which represents the estimated strength and direction of the association between the survival outcome (time to seizure freedom) and the predictor variables in the model. The "Association parameter" measures the association between the count data and zero-inflation, its 95% credible interval (0.121, 0.907) not includes zero, suggesting significant association.

**Diagnostic plots for joint model**

Similar to the survival and longitudinal mode in appendix (Appendix B, Figure B.4.) the density plot for joint were presented. The density plot was approximately similar to the normal plot. This result indicated that the joint model fitted the data well and it was converged.

**Comparison of Overall Performance of Models for separate and joint models**

To evaluate the overall performance of both the separate and joint models in terms of model parsimony and goodness of fit for the collected epilepsy patient's data at HUCSH. One criteria is which model was estimate the parameters with smaller standard errors, for this case from the joi nt model here for both responses in the survival and longitudinal sub-models significant parameters were estimated with a relatively smaller standard error. Two, the association parameter in the survival sub-model was highly statistical significance. So, all of the above two reasons gave an evidence that the joint model is productive and better fit than the separate models for this study.

**Discussion**

The main objective of this study was to identify of seizure attacks and time to seizure freedom in epileptic patients at Hawassa University Comprehensive specialized hospital. So, three different

models were explored, the generalized linear mixed effects model for average seizure attacks, parametric survival model for time to time to seizure freedom of epileptic patients separately and jointly modeling of the two outcomes together.

In current study out of the total epilepsy patients, 58.6% of males were seizures-free, compared to 31.0% of females who were seizures-free. This aligns with the findings of the (Abebe, K et.al, 2022) study where males had a higher rate of seizure freedom. The majority of government employees were seizures-free (56.2%), while among farmers, 15.8% were seizures-free. This finding is similar to the (Dubale, M et.al, 2020) study's results, which showed that government employees had a higher likelihood of being seizures-free. The education level seems to have a positive correlation with seizure freedom in both the second and third data sets. The more educated individuals (tertiary education) have higher rates of seizure freedom compared to those with lower education levels. Similar to the (Abebe, K et.al, 2022) study's findings, urban residents had a higher rate of seizure freedom (61.6%) compared to rural residents (27.6%). The current study's results align with the (Abebe, K et.al, 2022) study's finding that married patients had a higher rate of seizure freedom (58.6%).

The study conducted by (Dubale, M et.al, 2020) indicated that patients living with others had a higher rate of seizure freedom, consistent with the current study's finding that most patients live with others and have a higher rate of seizures-free status. Thus, the study's results showed a similar trend as the first study, where patients with higher economic statuses had a higher rate of seizure freedom. Also, the study's prevalence of generalized seizures (60%) is consistent with the (Abebe, K et.al, 2022) study's findings, where generalized seizures were the most common (98.3%). The current study reported that patients with no family history of epilepsy had a higher rate of seizure freedom, similar to the (Abebe, K et.al, 2022) study's results. The in our majority of patients received payment-based treatment 162(79.8%). similar to the first study where most patient medication sources were by payment (70.9%).

In this finding patients with brain injury epileptic compared to no brain injury patients were a worse prognosis of experiencing seizure freedom that not significantly associated with the survival model this is consistent with the finding done by (Hitiris et al., 2007), and Hitiris et al., (2007) reveals that mental retardation is not factors for seizure freedom this is inconsistent with this study.

The present study revealed that economic status of patients was not associated with the number o f seizure attacks which is different to study conducted by (Mekonnen, F. H et.al 2019) which is, l ower socioeconomic status of patients was associated with the number of seizure attacks and diff erent to (Hesdorffer DC et.al, 2005), the study suggested that low socio-economic status, indexed by low education or lack of home ownership, was a risk factor for epilepsy in adults. Other schol ars also pointed out that people with epilepsy in developing regions carry a heavy burden of stig ma associated with poor social and economic status (Baskind R et.al, 2005).

Patients who had took Alcohol have no statistical association with seizure freedom, this result is different to the study done in Amanuel Mental Specialized Hospital, Addis Ababa, Ethiopia by ( Waja et al., 2016) which is consumption of alcohol leads long-term hospital admission and poor seizure control, this might be neurotoxic effects of alcohol lead to epilepsy seizures and common ly perceived that patients with epilepsy experience problems with seizure control.

In this finding, having a family history of seizure were increased the seizure attacks but, the incre ase of follow-up time and believing to be cured decreased average seizure attacks. Moreover, ma rital status, educational background, employment status and place of residence were no significa nt factors of seizure attacks, this result agree to the study done by (Mekonnen et al., 2019), but co ntrary to these authors variables age and sex were not significant factors of this study. This differ ence might be methods of study, length of the study and sample size. But patients who had co-m orbidity less likely to experience seizure freedom compared to non-co-morbid patients this result is in line with the studies done by (Dessie et al., 2019, Hitiris et al., 2007).

# 5. CONCLUSION and RECOMMENDATION

## 5.1. Conclusions

This study was a five-year retrospective study based on 203 random samples of epilepsy patients who were attending their treatment at Hawassa University Comprehensive Specialized Hospital, between May 1, 2018, and May 1, 2023. For 203 heart failure patients, the data includes both a longitudinal interest in seizure episode and time to seizure free patients. The data were analyzed using a generalized linear mixed model for a separate count longitudinal model, Cox and Weibull parametric regression models for a separate survival model and a joint longitudinal-survival model using a Bayesian approach.

For Weibull survival model, variables such as Phenytoin usage, receiving more treatments, and having co-morbidities exhibited statistically significant associations with the hazard of achieving seizure freedom, while others like Phenobarbitone and Sodium Valproate did not show significant effects. For Zero-Inflated Negative Binomial (ZINB) model Phenytoin usage, receiving more treatments, exercise, alcohol consumption, having a family history of epilepsy, and the presence of co-morbidities were identified as factors significantly affecting the frequency of seizure attacks.

The joint modeling approach combining the Weibull survival model for time to seizure freedom and the ZINB model for seizure frequency showed promising results. Notably, variables related to Phenytoin usage, receiving more treatments, and having a comorbidity exhibited significant effects on both the time to seizure freedom and the frequency of seizure attacks.

This comprehensive Bayesian analysis provides valuable insights into the complex interplay of factors affecting seizure frequency and factors affecting survival time of seizure freedom among epileptic patients. This research contributes to a more nuanced understanding of epileptic patient management and could potentially guide personalized treatment strategies at Hawassa University Comprehensive Specialized Hospital.

## 5.2 Recommendation

Based on the Bayesian approach for joint modeling of time to seizure freedom and seizure frequency count in epileptic patients at Hawassa University Comprehensive Specialized

Hospital, the following recommendations has been made for both the hospital and the government:

- Given the statistically significant effects of certain antiepileptic drugs (AEDs) on seizure freedom and frequency, the hospital should consider tailoring treatment plans based on individual patient profiles.
- The presence of co-morbidities was found to be significantly associated with both seizure freedom and frequency. The hospital should prioritize comprehensive assessment and management of co-morbid conditions in epileptic patients to potentially improve their seizure control.
- The positive effect of exercise on reducing seizure frequency suggests that promoting and encouraging regular physical activity could be an effective adjunctive strategy for managing epilepsy. Hospital staff and healthcare providers should educate patients about the potential benefits of exercise and incorporate it into their treatment plans.
- While the study provides important insights, there are certain variables (such as stress, brain injury, and chronic diseases) that did not demonstrate statistically significant effects on seizure outcomes. Further research and larger sample sizes could provide more definitive conclusions regarding these variables' impact on epilepsy management.
- Overall, the findings of this study can guide Hawassa University Comprehensive Specialized Hospital in tailoring its treatment and management strategies for epileptic patients.

**REFERENCE**

Adal, Habtamu Digis, Kassahun Alemu, and Esileman Abdela Muche. "Seizure control status and associated factors among pediatric epileptic patients at a neurologic outpatient clinic in Ethiopia." Plos one 16.11 (2021): e0259079.

Allen, Andrew S., and Huiman X. Barnhart. "General marginal regression models for the joint modeling of event frequency and correlated severities with applications to clinical trials." Journal of Data Science 3.2 (2005): 199-219.

Allers, Katharina, et al. "The economic impact of epilepsy: a systematic review." BMC neurology 15 (2015): 1-16.

Baghfalaki, T., et al. "A Bayesian shared parameter model for joint modeling of longitudinal continuous and binary outcomes." Journal of Applied Statistics 49.3 (2022): 638-655.

Beghi, Ettore, et al. "Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016." The Lancet Neurology 18.4 (2019): 357-375.

Belayneh, Zelalem, and Birhanie Mekuriaw. "A systematic review and meta-analysis of anti-epileptic medication non-adherence among people with epilepsy in Ethiopia." Archives of Public Health 78 (2020): 1-14.

Birru, Eshetie Melese, Miftah Shafi, and Mestayet Geta. "Drug therapy of epileptic seizures among adult epileptic outpatients of University of Gondar Referral and Teaching Hospital, Gondar, North West Ethiopia." Neuropsychiatric disease and treatment (2016): 3213-3219.

Chen, Liddy M., Joseph G. Ibrahim, and Haitao Chu. "Sample size and power determination in joint modeling of longitudinal and survival data." Statistics in medicine 30.18 (2011): 2295-2309.

Cox, David R. "Regression models and life-tables." Journal of the Royal Statistical Society: Series B (Methodological) 34.2 (1972): 187-202.

Dessiso, Ahmed Hasan, and Ayele Taye Goshu. "Bayesian joint modelling of longitudinal and survival data of HIV/AIDS patients: a case study at bale robe general hospital, Ethiopia." American Journal of Theoretical and Applied Statistics 6.4 (2017): 182-190.

Dubale, Mariam, et al. "Treatment outcome and associated factors among adult epileptic patients at Hawassa University Specialized Hospital, Southern Ethiopia." J Bioanalysis Biomed 12.6 (2020): 7.

Duko, Bereket, et al. "Depressive Symptoms among Patients with Epilepsy Attending Adare General Hospital, Hawassa Ethiopia: Cross-sectional Study." (2018).

Fanta, Tolesa, et al. "Prevalence and factors associated with perceived stigma among patients with epilepsy in Ethiopia." Psychiatry journal 2015 (2015).

Fehintola, Funmito O., et al. "Knowledge of and attitude towards epilepsy among women in Ile-Ife, Nigeria." Sudan Journal of Medical Sciences 14.1 (2019): 31-44.

Fiest, Kirsten M., et al. "Prevalence and incidence of epilepsy: a systematic review and meta-analysis of international studies." Neurology 88.3 (2017): 296-303.

Fisher, Robert S., et al. "ILAE official report: a practical clinical definition of epilepsy." Epilepsia 55.4 (2014): 475-482.

Fite, Robera Olana, and Mistire Teshome Guta. "Stigma and associated factors among people with epilepsy in Ethiopia: A systematic review and meta-analysis." Epilepsy & Behavior 117 (2021): 107872.

Gebre, Abadi Kahsu, and Amdemicheal Haylay. "Sociodemographic, clinical variables, and quality of life in patients with epilepsy in Mekelle City, Northern Ethiopia." Behavioural neurology 2018 (2018).

Gebrewold, Meron Awraris, et al. "Ethiopian teachers: their knowledge, attitude and practice towards epilepsy." BMC neurology 16 (2016): 1-8.

Gillespie, Brenda. "Checking assumptions in the cox proportional hazards regression model." Midwest SAS Users Group (MWSUG) (2006).

Hassen, Oumer, and Ayalew Beyene. "The effect of seizure on school attendance among children with epilepsy: a follow-up study at the pediatrics neurology clinic, Tikur Anbessa specialized hospital, Addis Ababa, Ethiopia." BMC pediatrics 20.1 (2020): 1-7.

Henderson, Robin, Peter Diggle, and Angela Dobson. "Joint modelling of longitudinal measurements and event time data." Biostatistics 1.4 (2000): 465-480.

Hickey, Graeme L., et al. "A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial." Journal of the Royal Statistical Society Series A: Statistics in Society 181.4 (2018): 1105-1123.

Hickey, Graeme L., et al. "A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial." Journal of the Royal Statistical Society. Series A (Statistics in Society) 181.4 (2018): 1105-1123.

Hu, Yin, et al. "Gender and socioeconomic disparities in global burden of epilepsy: an analysis of time trends from 1990 to 2017." Frontiers in Neurology 12 (2021): 643450.

Karasoy, Duru. "Joint modeling of a longitudinal measurement and parametric survival data with application to primary biliary cirrhosis study." Pakistan Journal of Statistics and Operation Research 16.2 (2020).

Kim, Lois G., et al. "Prediction of risk of seizure recurrence after a single seizure and early epilepsy: further results from the MESS trial." The Lancet Neurology 5.4 (2006): 317-322.

Kwan, Patrick, et al. "Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies." (2010): 1069-1077.

Lee, Keunbaik, and Yongsung Joo. "Marginalized models for longitudinal count data." Computational Statistics & Data Analysis 136 (2019): 47-58.

Martino, Sara, and Håvard Rue. "Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the inla program." Department of Mathematical Sciences, NTNU, Norway (2009).

Mekonnen, Fenta Haile, et al. "Statistical models for longitudinal zero-inflated count data: application to seizure attacks." African Health Sciences 19.3 (2019): 2555-2564.

Melkamu, Piniel, et al. "Cost of illness of epilepsy and associated factors in patients attending adult outpatient department of University of Gondar Referral Hospital, Northwest Ethiopia." Risk Management and Healthcare Policy (2021): 2385-2394.

Minwuyelet, Fentahun, et al. "Quality of life and associated factors among patients with epilepsy at specialized hospitals, Northwest Ethiopia; 2019." Plos one 17.1 (2022): e0262814.

Raru, Temam Beshir, Bisrat Misganaw Geremew, and Koku Sisay Tamirat. "Change in the frequency of seizure attacks and associated factors among adult epilepsy patients at Amanuel Mental Specialized Hospital (AMSH): a generalized linear mixed model (GLMM)." Neuropsychiatric Disease and Treatment (2021): 2529-2538.

Raru, Temam Beshir, Bisrat Misganaw Geremew, and Koku Sisay Tamirat. "Change in the frequency of seizure attacks and associated factors among adult epilepsy patients at amanuel mental specialized hospital (Amsh): A generalized linear mixed model (glmm)." Neuropsychiatric Disease and Treatment (2021): 2529-2538.

Rizopoulos, Dimitris. Joint models for longitudinal and time-to-event data: With applications in R. CRC press, 2012.

Rizopoulos, Dimitris. Joint models for longitudinal and time-to-event data: With applications in R. CRC press, 2012.

Rogers, J. K., and J. L. Hutton. "Joint modelling of pre-randomisation event counts and multiple post-randomisation survival times with cure rates: application to data for early epilepsy and single seizures." Journal of Applied Statistics 40.3 (2013): 546-562.

Shegaye, Shumet, et al. "Antiepileptic Drug Adherence and Its Associated Factors among Epilepsy Patients on Follow-ups at Emanuel Mental Specialized Hospital, Ethiopia." Ethiopian Journal of Health Sciences (2022): 913-922.

Shuma, Firafan, Behailu Terefe, and Tamirat Tekassa. "Epilepsy Treatment Outcome, Adherence to Anti-seizure Medications and Predicting factors at the chronic care facility in Jimma Universtiy Medical Center, Jimma, Southwest Ethiopia: Cross-sectional study." (2022).

Singh, Gagandeep, and Josemir W. Sander. "The global burden of epilepsy report: Implications for low-and middle-income countries." Epilepsy & Behavior 105 (2020): 106949.

Sousa, Inês. "A Review on Joint Modelling of Longitudinal Measurements and Time-To-Even." REVSTAT-Statistical Journal 9.1 (2011): 57-81.

Sunethra, A., and R. Sooriyarachchi. "A joint model for exponential survival data and Poisson count data." American Journal of Applied Mathematics and Statistics 6.2 (2018): 72-9.

Thijs, Roland D., et al. "Epilepsy in adults." The Lancet 393.10172 (2019): 689-701.

Tigistu, Mekdes, et al. "Frequency of seizure attack and associated factors among patients with epilepsy at University of Gondar Referral Hospital: a cross-sectional study, Gondar, North West Ethiopia, 2017." BMC research notes 11.1 (2018): 1-6.

Twisk, Jos WR. Applied longitudinal data analysis for epidemiology: a practical guide. cambridge university press, 2013.

Vaid, Nidhi, et al. "Epilepsy, poverty and early under-nutrition in rural Ethiopia." Seizure 21.9 (2012): 734-739.

Waja, Tsegereda, et al. "Prevalence of alcohol use disorders and associated factors among people with epilepsy attending Amanuel Mental Specialized Hospital, Addis Ababa, Ethiopia." Neuropsychiatric disease and treatment (2016): 2989-2996.

WHO Fact sheet: Epilepsy. Retrieved from https://www.who.int/news-room/fact-sheets/detail/epilepsy (2019).

Wubetu, Abate Dargie, Elyas Admasu Basha, and Nigus Alemnew Engidaw. "Public knowledge and attitude towards epilepsy and its associated factors: community-based cross-sectional study, Ethiopia, 2019." Journal of Environmental and Public Health 2020 (2020).

Xian, L. I. U. "Introduction to longitudinal data analysis in psychiatric research." Shanghai archives of psychiatry 27.4 (2015): 256.

Yazie, Taklo Simeneh, Belayneh Kefale, and Mulugeta Molla. "Treatment outcome of epileptic patients receiving antiepileptic drugs in Ethiopia: a systematic review and meta-analysis." Behavioural neurology 2021 (2021).

Zena, Dawit, et al. "Seizure control and its associated factors among epileptic patients at Neurology Clinic, University of Gondar hospital, Northwest Ethiopia." SAGE Open Medicine 10 (2022): 20503121221100612.

Zhang, Fan, et al. "Assessing importance of biomarkers: A Bayesian joint modelling approach of longitudinal and survival data with semi-competing risks." Statistical modelling 21.1-2 (2021): 30-55.

## APPENDIX

### APPENDIX A: Individual profile plots for categorical variables

*Figure 1. Individual profile plots for types of AED*

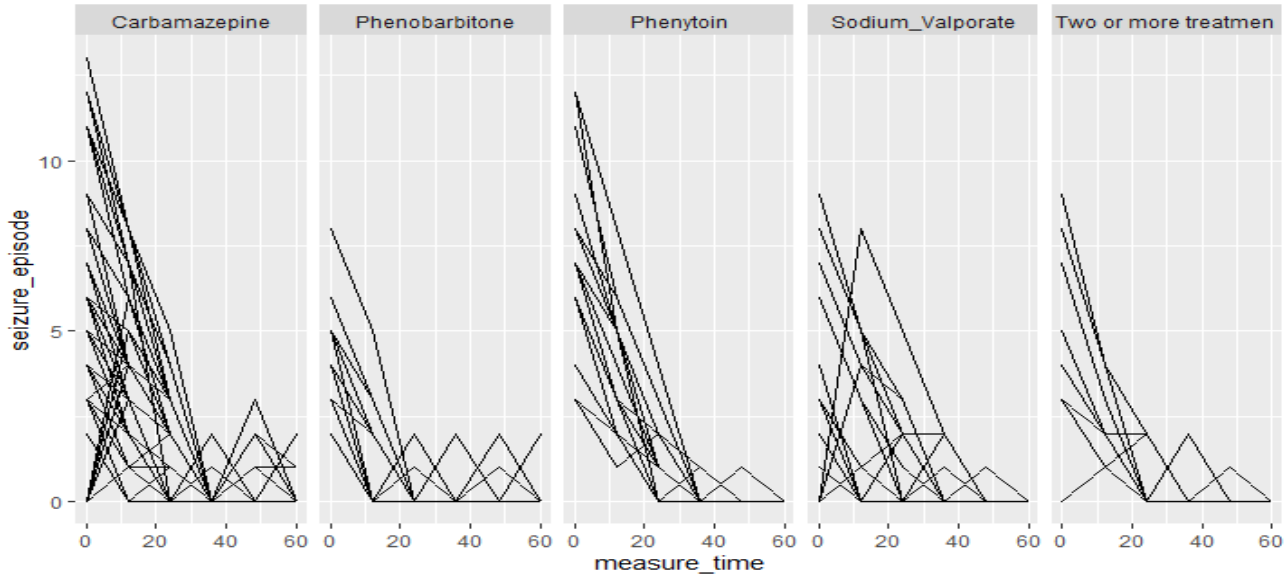



*Figure 2. Individual profile plots for types of type_seiz*

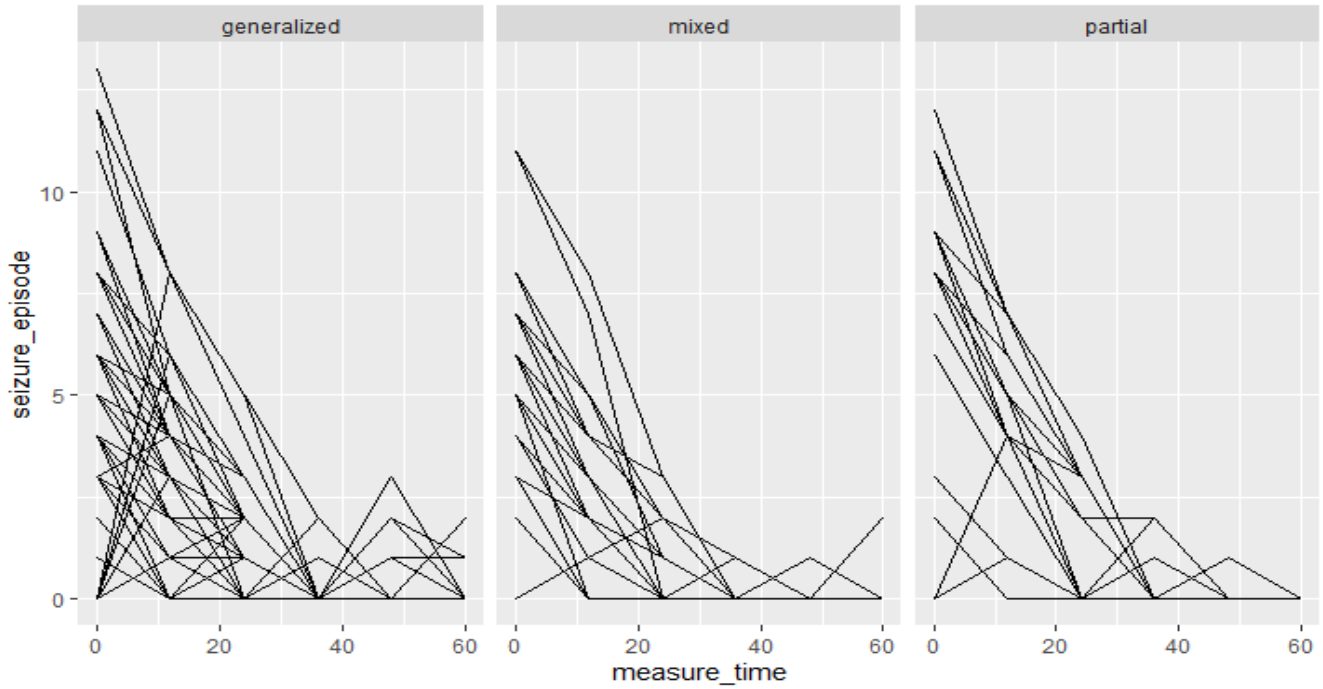


*Figure 3. Individual profile plots for types of exercise*

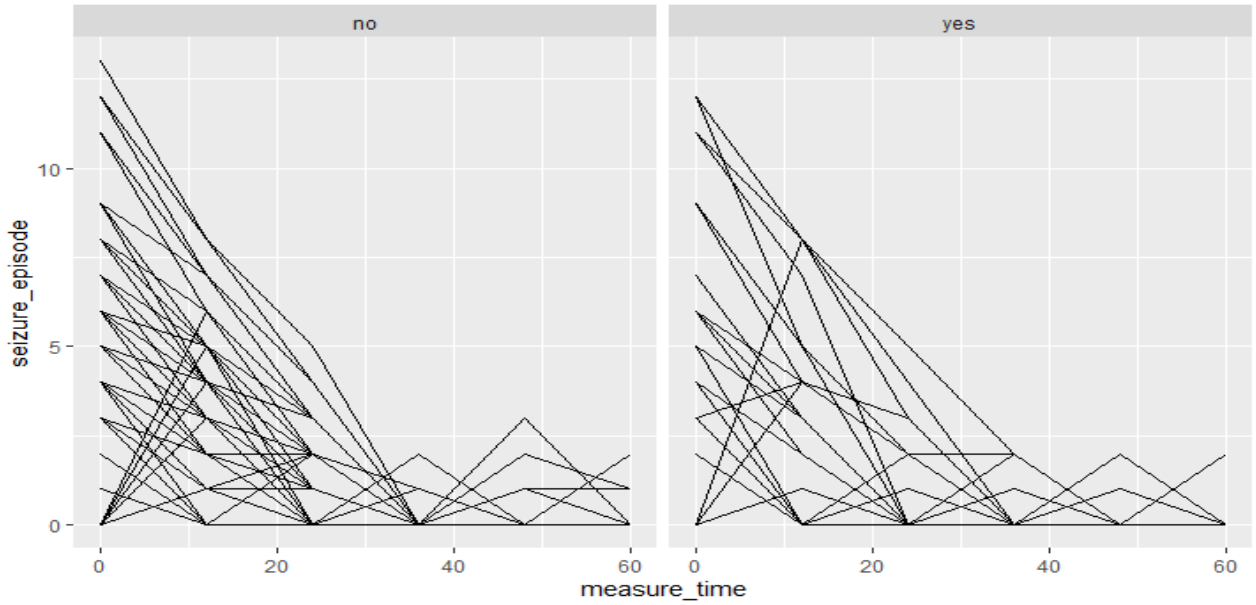



*Figure 4. Individual profile plots for types of co morbidity*

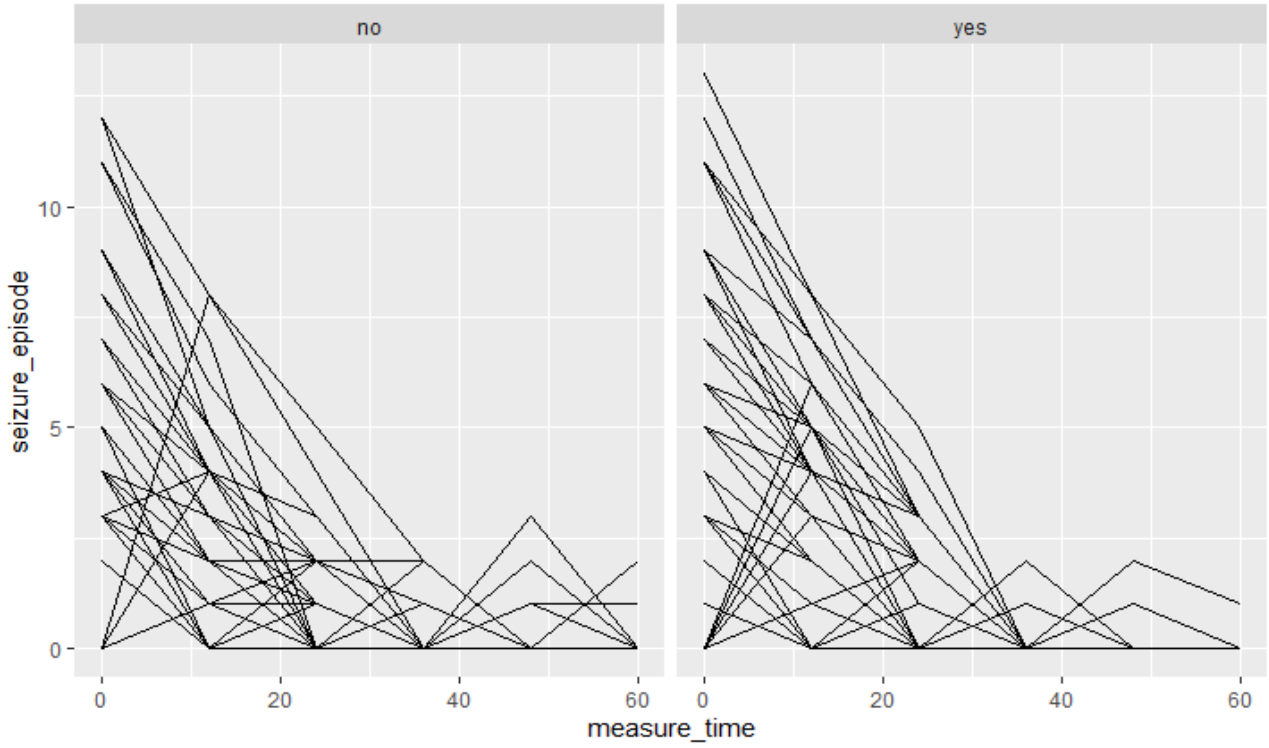



*Figure 5. Individual profile plots for types of family history*

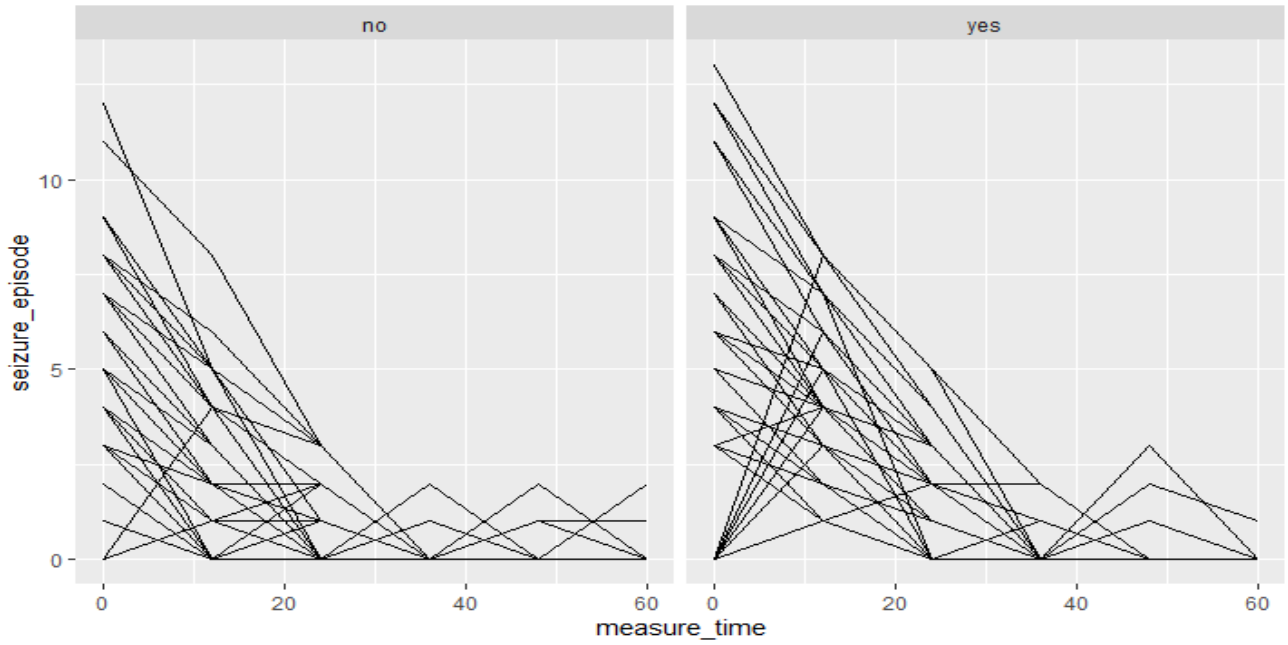



*Figure 6. Individual profile plots for types of brain injury*

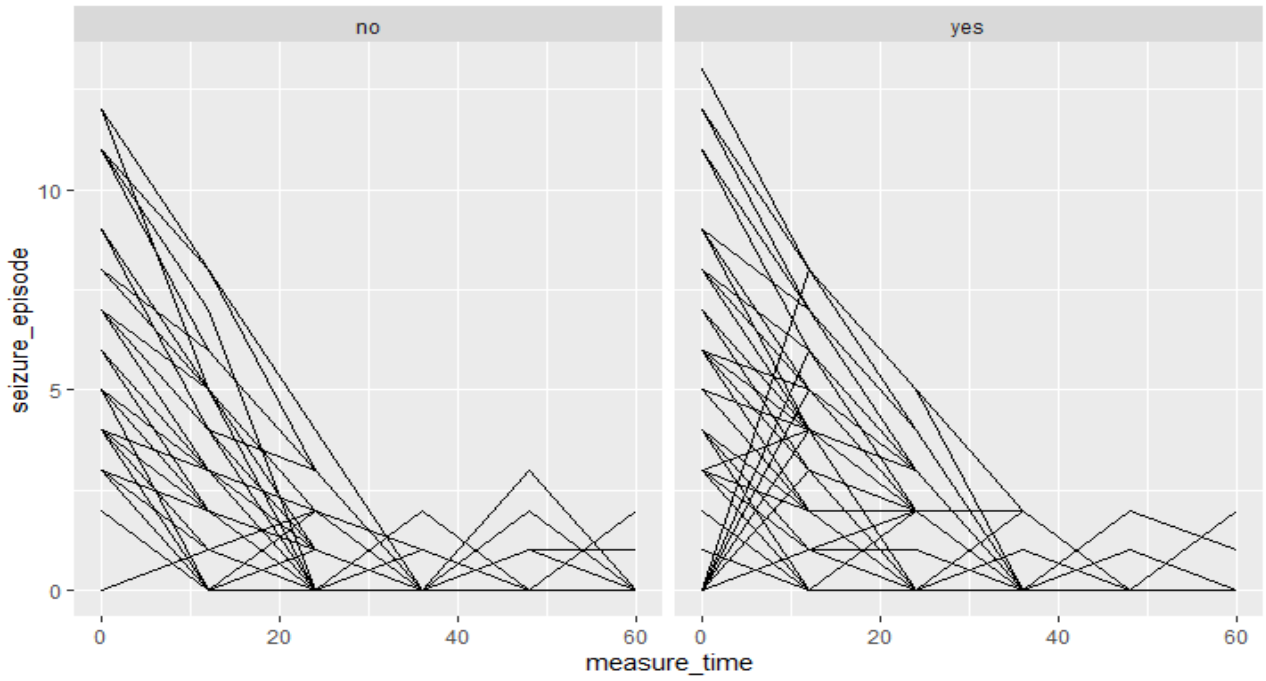



*Figure 7. Individual profile plots for types of alcohol*

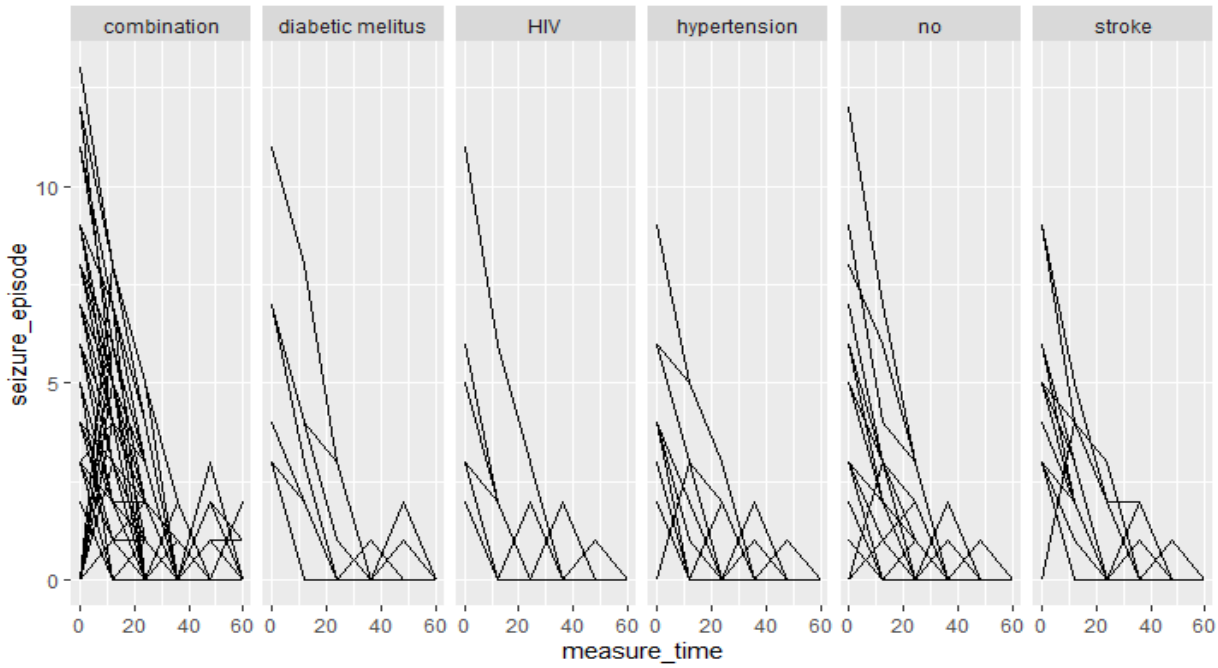
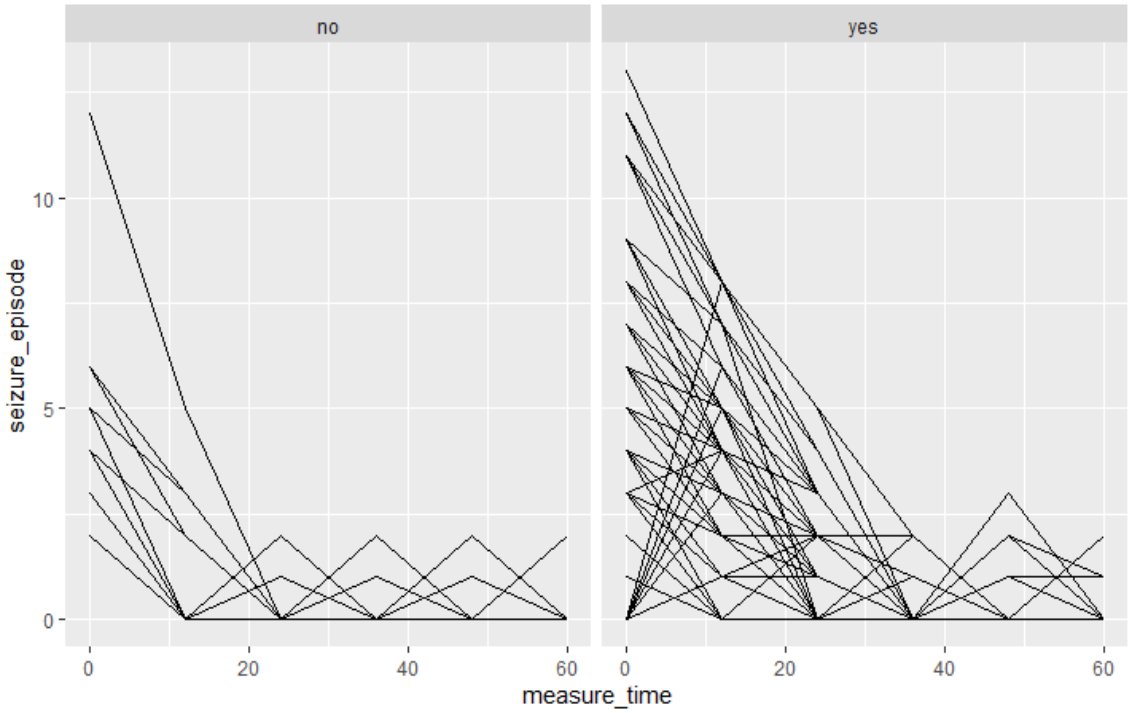



*Figure 8. Individual profile plots for types of chronic disease*



**APPENDIX B: Diagnostic plots**

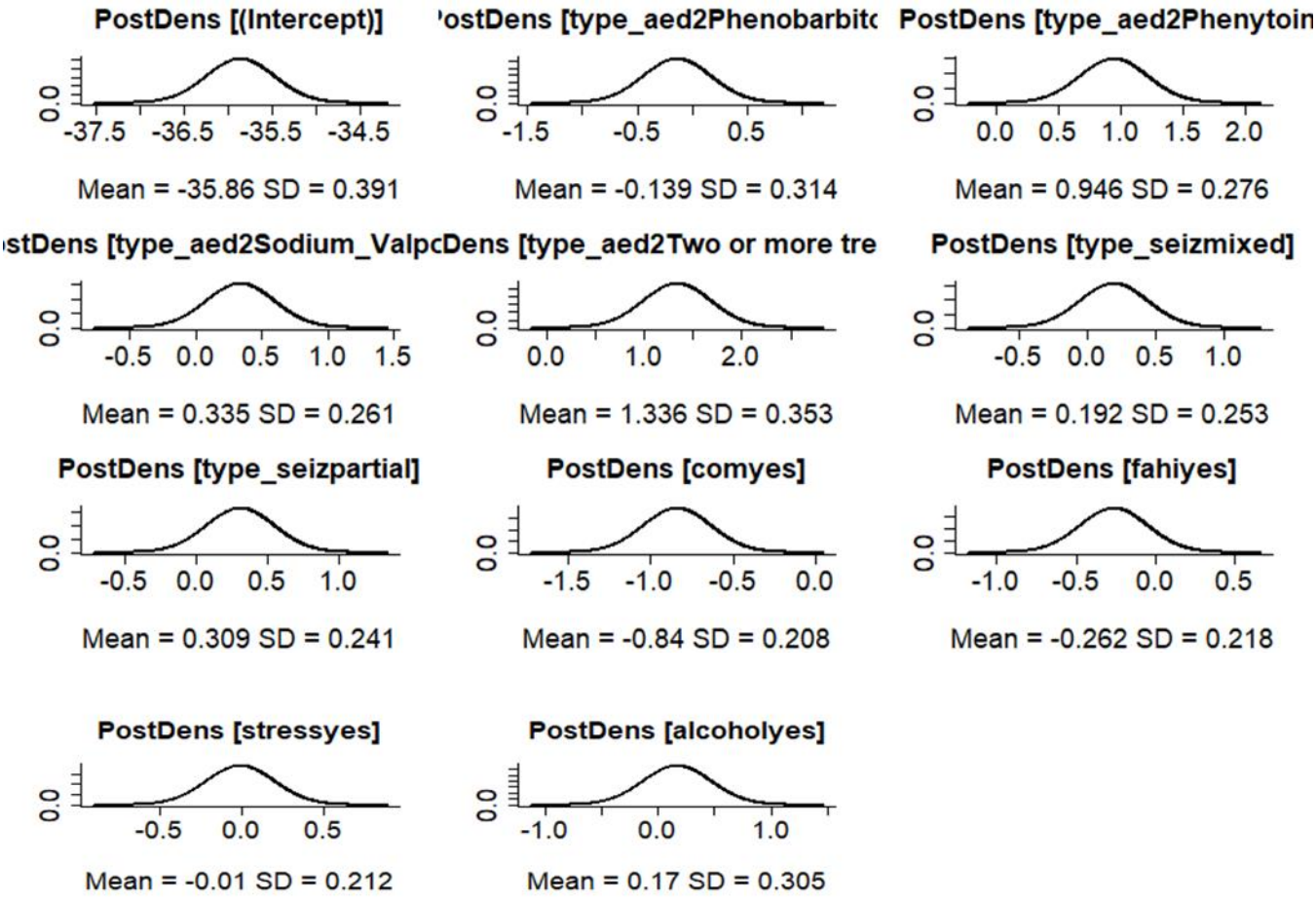



Figure B.1. Estimated density plots for the predictors of survival model

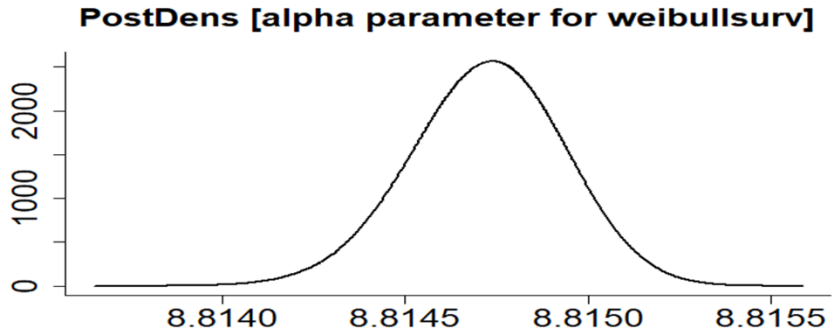



Figure B.2. Estimated density plots for the baseline hazard of Weibull survival model

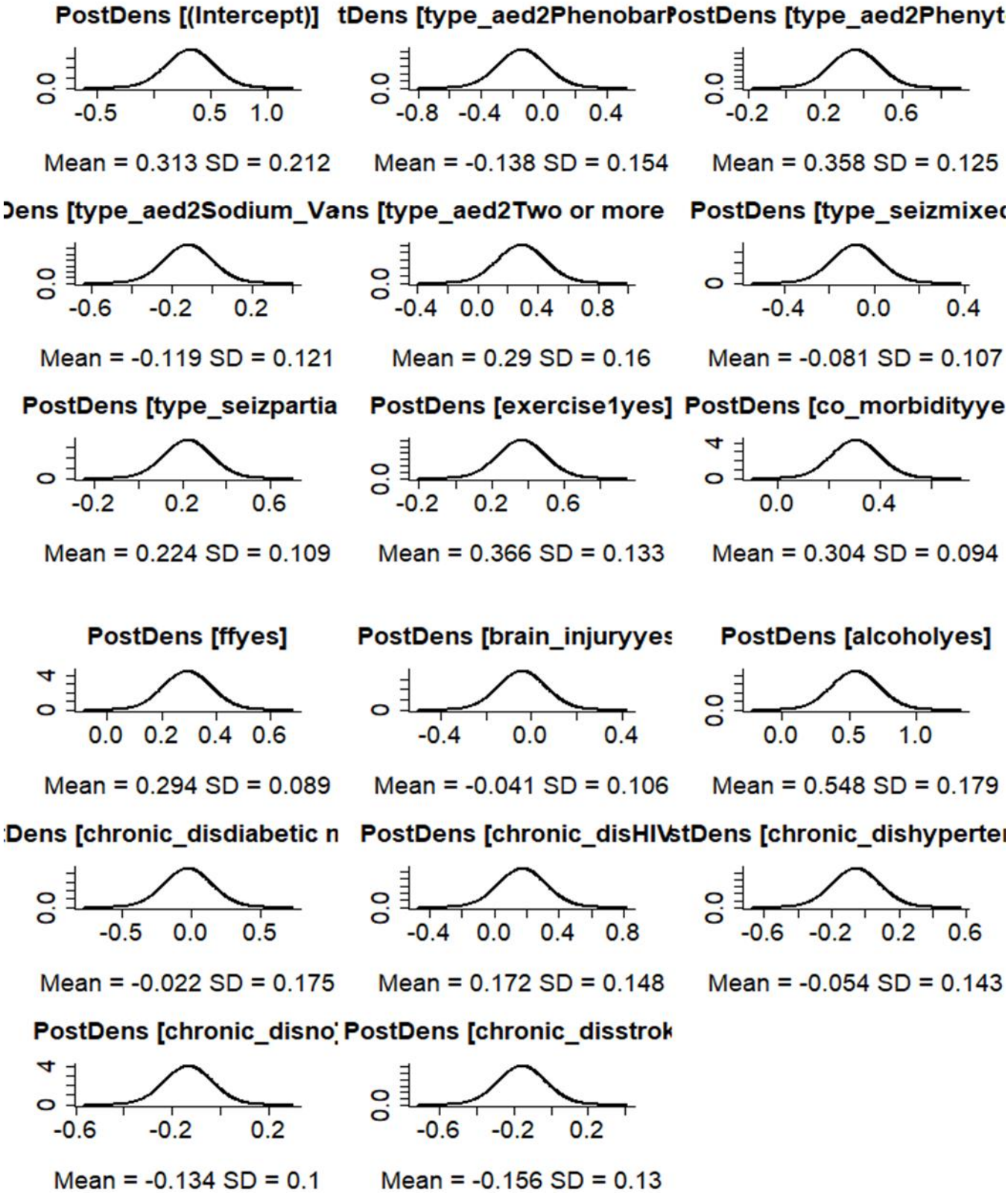
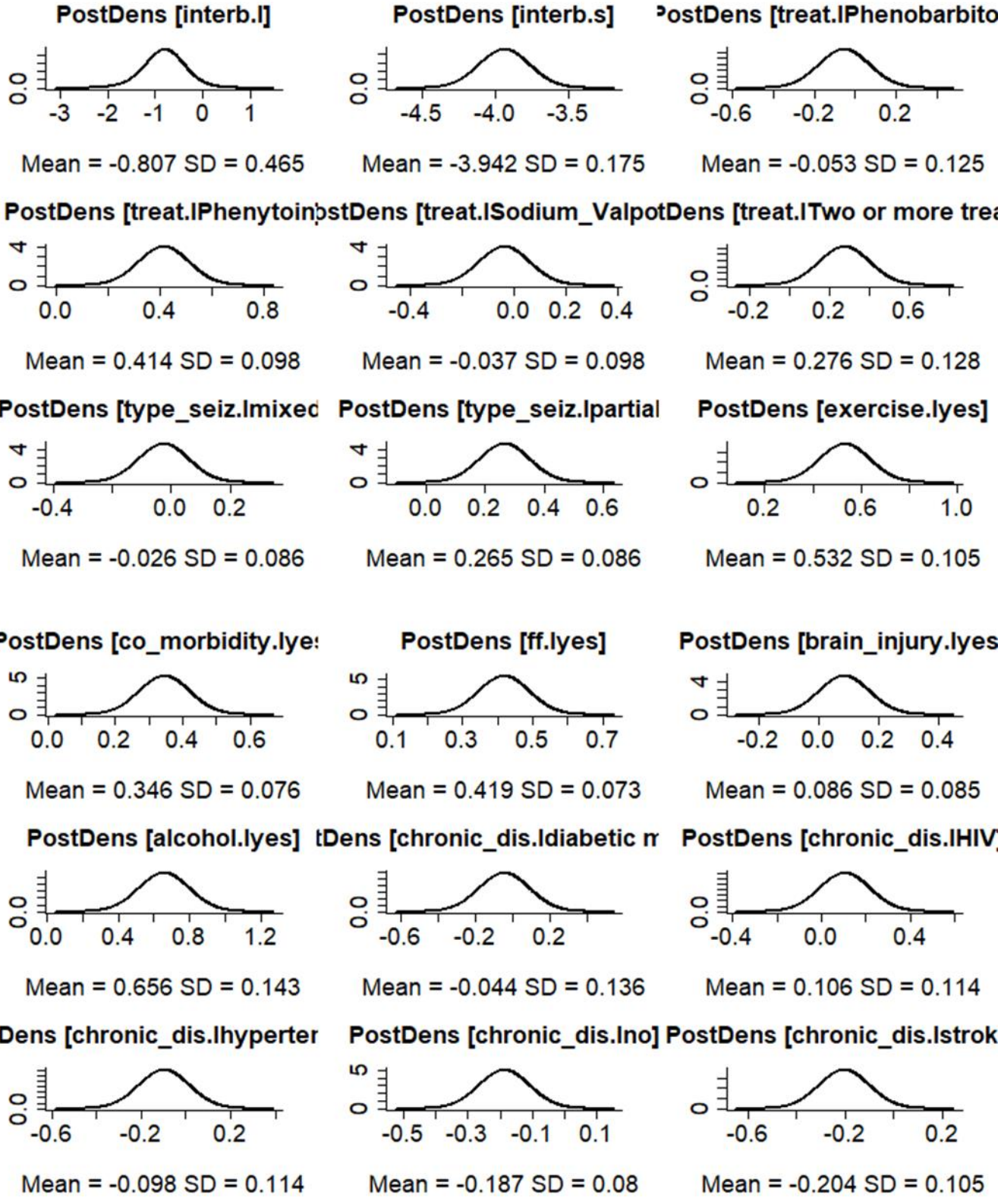



Figure B.3. Estimated density plots for the predictors of longitudinal model


PostDens [interb.l]
Mean = -0.807 SD = 0.465
PostDens [interb.s]
Mean = -3.942 SD = 0.175
PostDens [treat.lPhenobarbito
Mean = -0.053 SD = 0.125
PostDens [treat.lPhenytoin
Mean = 0.414 SD = 0.098
stDens [treat.lSodium_Valpot
Mean = -0.037 SD = 0.098
Dens [treat.lTwo or more trea
Mean = 0.276 SD = 0.128
PostDens [type_seiz.lmixed
Mean = -0.026 SD = 0.086
PostDens [type_seiz.lpartial
Mean = 0.265 SD = 0.086
PostDens [exercise.lyes]
Mean = 0.532 SD = 0.105
PostDens [co_morbidity.lye
Mean = 0.346 SD = 0.076
PostDens [ff.lyes]
Mean = 0.419 SD = 0.073
PostDens [brain_injury.lyes
Mean = 0.086 SD = 0.085
PostDens [alcohol.lyes]
Mean = 0.656 SD = 0.143
tDens [chronic_dis.ldiabetic n
Mean = -0.044 SD = 0.136
PostDens [chronic_dis.lHIV
Mean = 0.106 SD = 0.114
stDens [chronic_dis.lhyperter
Mean = -0.098 SD = 0.114
PostDens [chronic_dis.lno]
Mean = -0.187 SD = 0.08
PostDens [chronic_dis.lstrok
Mean = -0.204 SD = 0.105

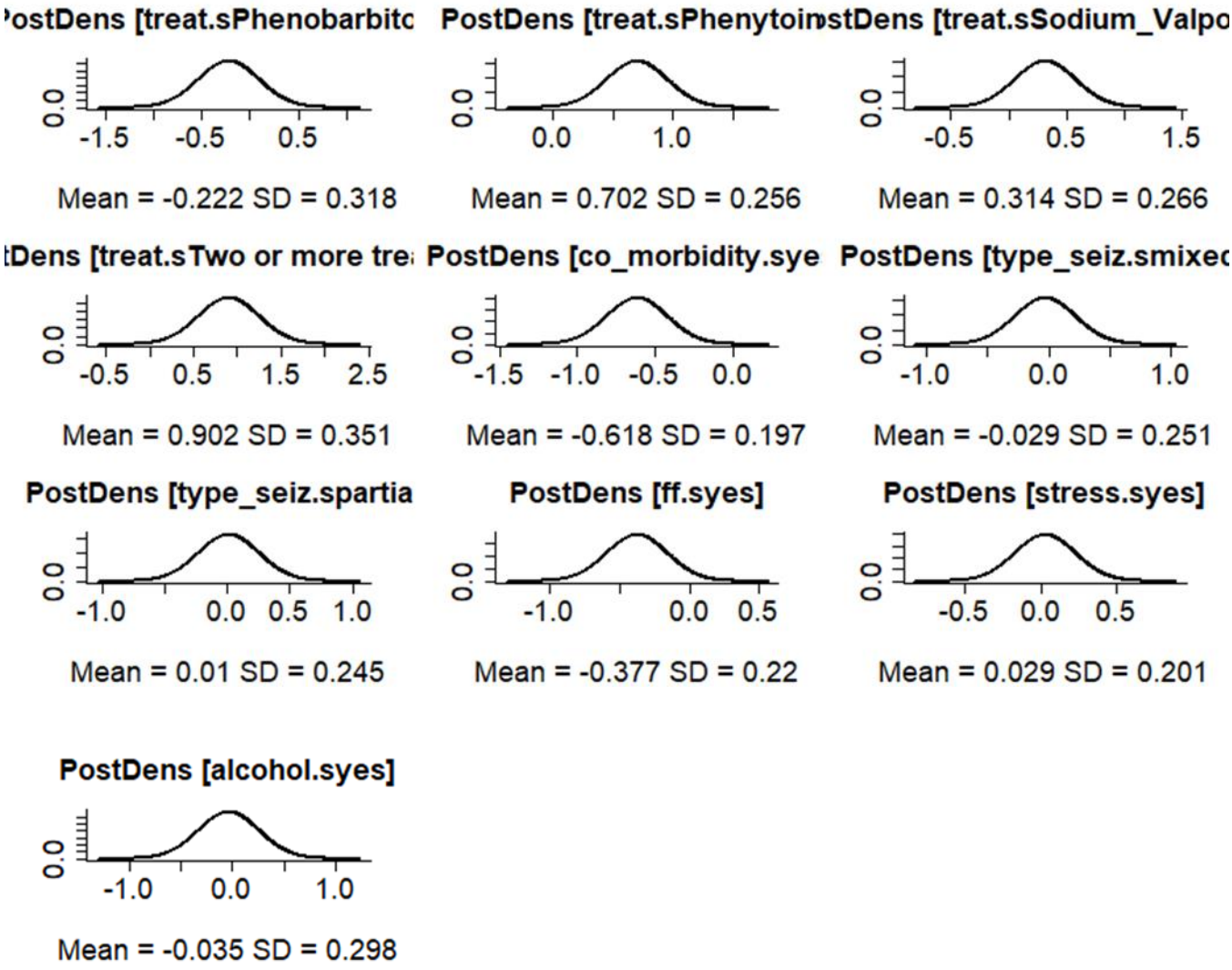



Figure B.4. Estimated density plots for the predictors of longitudinal model

**APPENDIX C. Questionnaire**

This interview is prepared for the purpose of eliciting information on the study titled determinant of seizure attacks and seizure freedom for epilepsy patients at HUCSH. Using Separate and Joint Model Application. The information you provide will be used only for research purpose and you are guaranteed that the researcher has no any other hidden agenda. I would like to thank you in advance for your cooperation and patience in the interview session since your participation is invaluable to the success of this study.

1. Medication identification card number_____

2. Duration of taking treatment in this hospital_____

3. Sex of the respondent  male □  female □

4. Your age now_____

5. Residence  rural □ urban □

6. Religion orthodox □  Muslim □ protestant □ other □

7. Marital status:  single □ married □ divorced □ widowed □

8. Living status: alone □ with others □

9. Education level illiterate □ religious school □ primary □ secondary □ tertiary □

10. Occupation: government employee □  student □ farmer □ skill labor □ no work

11. Economic status rich □ medium □ low □

12. Did you do an exercise/play?  Yes □ no □

13. Do you have a first and second family history of epilepsy seizure?  Yes □ no □

14. Epilepsy syndrome / Cause  cryptogenic □ idiopathic □ symptomatic □

15. Do you have a mental disability/retardation?  Yes □ no □

16. Age at seizure onset_____

17. History of febrile convulsions Yes □ no □

18. History of neonatal seizure Yes □ no □

19. Drug type_____

20. The number of drug used still now_____

21. Type of seizures generalized □ partial □

22. Did you take epilepsy treatment before starting this hospital? Yes □ no □

23. Do have another disease in addition to epilepsy seizure Yes □ no □

24. If you said yes in Q23 what type of disease? _____

25. Was there a side effect of drugs after taking? Yes □ no □

26. Do you have a sleep deprivation? Yes □ no □

27. Do you take a meal on time? Yes □ no □

28. Do you think people discriminate you? Yes □ no □

29. Do you believe modern AEDs treats you rather than tradition/other? Yes □ no □

30. Do you believe contagion? Yes □ no □

31. Do you believe you will be cured? Yes □ no □

32. Do have stress? Yes □ no □

33. Number of seizures in the past visit_____.

34. Do you satisfied with this hospital service? Yes □ no □

35. Many of your regular follow ups were? Per month □ 2months □ three months □ >3months□

36. Are drug addicted? yes □ no □ if yes alcohol □ chat □ others □

37. Your birth interval_____

38. Your family size_____

39. Do you get a family support? Yes □ no □

40. Head trauma/brain injury Yes □ no □

41. Source of medication free □ payment □

42. Have experience chronic diseases (Diabetic mellitus) Yes □ no □

43. Have experience chronic diseases (Hypertension) Yes □ no □

44. Do you take traditional drugs before started to diagnosis?

**Assessment of adherence**

1. Do you sometimes forget to take your pills? Yes □ no □

2. People sometimes miss taking their medications for reasons other than forgetting. Thinking over the past two weeks, were there any days when you did not take your medicine?

Yes □ no □

3. Have you ever cut back or stopped taking your medicine without telling your doctor because you felt worse when you took it? Yes □ no □

4. When you travel or leave home, do you sometimes forget to bring along your medicine?

5. Did you take all your medicine yesterday? Yes □ no □

6. When you feel like your symptoms are under control, do you sometimes stop taking your medicine? Yes □ no □

7. Taking medicine every day is a real inconvenience for some people. Do you ever feel hassled about sticking to your treatment plan? Yes □ no □

8. How often do you have difficulty remembering to take all your medicine? Never □ once in a while sometimes □ usually □ all the time □