Modelling Child Survival In Malaria-Endemic Regions Of Kenya Using Bayesian Generalized Log-Logistic Models

Masyuko, Baraka Ngumbau

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Statistical Science of Strathmore University

Strathmore Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

June 2025

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the thesis itself.

Name: ..... **Masyuko Baraka Ngumbau** .....

Student Signature:

Date: ..... May 24, 2025 .....

Approval

The thesis of Masyuko Baraka Ngumbau was reviewed and approved for examination by the following:

Prof. Samuel Mwalili

Supervisor,

Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu

Dean,

Institute of Mathematical Sciences, Strathmore University.

Prof. Bernard Shibwabo

Director,

Office of Graduate Studies, Strathmore University.

Abstract

Child mortality remains a major public health challenge in malaria-endemic regions, particularly in sub-Saharan Africa, where children under five face disproportionately high risks. In Kenya, the burden is amplified by factors such as limited healthcare access, socioeconomic disparities, and insufficient malaria prevention. This study develops and applies a Bayesian Generalized Log-Logistic (GLL) survival model to examine how maternal health practices, regional and household-level socioeconomic conditions influence child survival in malaria-endemic regions. Using nationally representative data from the 2020 Kenya Demographic and Health Survey (KDHS) and Malaria Indicator Survey (MIS), the model captures a wide range of hazard shapes, including both increasing and decreasing risk patterns. To strengthen the model's theoretical foundation, its asymptotic properties were derived to ensure consistency and efficiency in parameter estimation under large-sample conditions. The Bayesian framework further enables robust inference by incorporating prior information and quantifying uncertainty. Posterior predictive checks demonstrated good model fit, confirming the model's capacity to reflect the observed survival dynamics. Key predictors of child survival included antenatal care utilization, household wealth, regional malaria endemicity, and malaria prevention behaviors. The study concludes that the Bayesian GLL model is a robust and flexible tool for understanding child mortality risk and can inform the design of targeted public health interventions in high-burden settings like Kenya.

Table of contents

Abstract iii

List of figures vii

List of tables viii

List of abbreviations ix

Acknowledgement x

Dedication xi

1 Introduction 1

1.1 Background to the study 1

1.2 Problem Statement 2

1.3 Objectives 3

1.3.1 General Objectives 3

1.3.2 Specific Objectives 3

1.4 Scope of the Study 3

1.5 Significance of the Study 4

2 Literature Review 5

2.1 Introduction 5

2.2 Malaria and Child Survival 5

2.3 Socioeconomic, Regional Disparities and Maternal Health Factors in Child Survival 6

2.4 Survival Modeling Techniques in Child Mortality Studies 7
2.5 Bayesian Inference in Survival Analysis 9

3 Methodology **14**
3.1 Introduction 14
3.2 Source of Data 14
3.3 Data Preprocessing 15
3.3.1 Handling Missing Values 15
3.4 Model Description 15
3.4.1 Survival Function 16
3.4.2 Likelihood Function 16
3.5 Bayesian Inference for the GLL Model 17
3.5.1 Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS) 21
3.5.2 Hamiltonian Monte Carlo (HMC) 21
3.5.3 No-U-Turn Sampler (NUTS) 22
3.5.4 Implementation in Stan 23
3.5.5 Posterior Mode Estimation 26
3.5.6 Posterior Predictive Validation 35
3.5.7 Ethical Considerations 35

4 Results and Interpretation **36**
4.1 Introduction 36
4.2 Frequency distribution 36
4.3 Explanatory Data Analysis 39
4.3.1 Missing Data Analysis 39
4.3.2 Statistical Analysis Results 42
4.4 Bayesian Hierarchical GLL Model Results 46
4.4.1 Introduction 46
4.4.2 Justification for Bayesian Generalized Log-Logistic Model using Log-Cumulative Hazard Plot 47
4.4.3 Posterior distributions 54

4.4.4 Posterior Estimates of Model Parameters 55
4.4.5 Population Attributable Fraction (PAF) Analysis 59
4.4.6 Analysis of Posterior Estimates of Regional Effects 62
4.5 Model Validation and diagnostics 63
4.5.1 Trace plots 63
4.5.2 Density Overlay Plot for Posterior Predictive Check 65
4.5.3 Posterior predictive uncertainty plot 66
4.5.4 Observed vs. Posterior Predictive Summary Statistics 67

5 Discussions, Conclusions and Recommendations 69
5.1 Introduction 69
5.2 Summary of Key Findings 69
5.3 Implications of Findings 70
5.3.1 Public Health and Policy Implications 70
5.3.2 Statistical and Methodological Contributions 71
5.4 Limitations of the Study 72
5.5 Recommendations for Future Research 72
5.6 Conclusion 73

References 74

Appendix A Similarity Index 78

Appendix B Ethical Clearance Confirmation 80

Appendix C R code 81

List of figures

Figure 4.1: Distribution of Child Survival Time Across Malaria Endemicity Zones in Kenya 45
Figure 4.2: Log-Cumulative Hazard Plot for Assessing Proportional Hazards Assumption 47
Figure 4.3: Posterior distributions for covariates 55
Figure 4.4: Posterior estimates of regional effects from the Bayesian Generalized Log-Logistic (GLL) survival model. 63
Figure 4.5: Trace Plots for parameters 64
Figure 4.6: Trace Plots for parameters 65
Figure 4.7: Posterior predictive check 66
Figure 4.8: Posterior Predictive Uncertainty Plot 67

List of tables

Table 4.1: Frequency Distribution of Key Variables 37

Table 4.2: Missing Values Summary 40

Table 4.3: Statistical Analysis Results 42

Table 4.4: Summary of Statistical Tests and Key Findings 43

Table 4.5: Adjusted Residuals from Chi-Square Test of Wealth and Malaria Endemicity 44

Table 4.6: Ordinal Regression Model for Predicting Wealth Index by Malaria Zone 45

Table 4.7: Justification of Prior Choices Based on Literature 54

Table 4.8: Posterior Estimates of Model Parameters 56

Table 4.9: Hazard Ratios and Effect Sizes for Risk Factors 59

Table 4.10: Population Attributable Fractions (PAFs) for Risk Factors 61

Table 4.11: Observed vs. Posterior Predictive Summary Statistics 67

List of abbreviations

GLL	Generalized Log-Logistic
HR	Hazard Ratio
MoH	Ministry of Health
WHO	World Health Organization

Acknowledgement

I thank the Almighty God for the knowledge, strength and good health during the study period. I wish to extend this gratitude to my supervisor, Prof. Mwalili for his exceptional guidance, patience, and invaluable feedback, which have been instrumental in shaping this work. His expertise and availability have made this study both enriching and successful. I am profoundly grateful to my family—my parents and sister for their unwavering support, encouragement, and belief in my capabilities.

Dedication

This thesis is dedicated to God Almighty for giving me wisdom and good health.

Chapter 1

Introduction

1.1 Background to the study

Child mortality is a severe public health issue in malaria-endemic regions, particularly in sub-Saharan Africa, where malaria is a leading cause of death for children under five. The World Health Organization (WHO) highlights that malaria continues to be a major cause of death, particularly among children under five in sub-Saharan Africa. In 2015, an estimated 292,000 children in Africa died due to malaria, and in 2022, the WHO reported 608,000 malaria deaths globally, with the African region accounting for the majority of these fatalities (World Health Organization (2023)). In Kenya, malaria transmission is notably high, contributing significantly to mortality rates despite efforts to curb the disease through interventions such as insecticide-treated nets (ITNs) and antimalarial treatments. Children under 5 are particularly susceptible to malaria illness, infection and death. In 2015, malaria killed an estimated 303 000 under-fives globally, including 292 000 in the African Region. Between 2010 and 2015, the malaria mortality rate among children under 5 fell by an estimated 35%. Nevertheless, malaria remains a major killer of under fives, claiming the life of 1 child every 2 minutes (Organization (2016)).

Numerous studies have examined the role of socioeconomic factors, healthcare accessibility, and maternal health practices on child mortality. For instance, research in Western Kenya has demonstrated how factors like distance to health facilities and socioeconomic status affect child survival rates, with closer proximity to healthcare facilities correlating with reduced mortality risk. Studies also highlight the limitations of traditional survival models, like the Cox proportional hazards model, in capturing the complex and non-monotonic mortality risk patterns associated with malaria in children. Conventional survival models,

like the Cox proportional hazards model, are commonly used in public health but are limited in their flexibility to model non-monotonic hazards, which often characterize malaria's impact on child survival. Recent research suggests that more adaptable models, such as the Generalized Log-Logistic (GLL) model, could capture complex hazard patterns more accurately, especially when coupled with Bayesian inference to incorporate prior knowledge and regional-specific data (Gething, Peter W. and Patil, Anand P. and Smith, David L. (2017)).

1.2 Problem Statement

Child mortality in malaria-endemic regions like Kenya is influenced by dynamic factors such as seasonal malaria transmission, socioeconomic disparities, and maternal health practices. While the Cox proportional hazards model is widely used in survival analysis, its assumption of constant hazard ratios over time limits its ability to capture the complex, time-varying risks characteristic of malaria-related mortality. Although extensions of the Cox model, for example, time-dependent covariates or stratified models, can address some of these limitations, they often lack the flexibility to model non-linear hazard shapes or incorporate hierarchical structures for regional variability.

Alternative parametric models, such as the Weibull or log-logistic distributions, offer more flexibility but are constrained to monotonic hazards. The Generalized Log-Logistic (GLL) model, however, can accommodate both monotonic and non-monotonic hazard patterns, making it better suited for modeling child survival in settings where risks fluctuate due to malaria seasonality, healthcare access, or intervention timing. Furthermore, Bayesian inference enhances this framework by allowing the integration of prior knowledge and robust uncertainty quantification, which is critical for policy decisions in resource-limited settings.

This study addresses these gaps by proposing a Bayesian hierarchical GLL model tailored to child survival data in Kenya. The model not only captures complex hazard patterns but also evaluates the interplay of maternal, socioeconomic, and regional factors, providing a comprehensive tool for public health planning and intervention targeting.

1.3 Objectives

This research consists of a main objective and specific objectives as outlined below:

1.3.1 General Objectives

To develop and validate a Bayesian Generalized Log-Logistic (GLL) survival model that flexibly captures diverse hazard patterns (increasing, decreasing, and U-shaped) in child survival data within malaria-endemic regions in Kenya and quantifying the effects of maternal health and socioeconomic factors.

1.3.2 Specific Objectives

The specific objectives for this study are:

i. To develop a Bayesian Generalized Log-Logistic (GLL) survival model to capture both increasing, decreasing and U-shaped hazard patterns in child survival data within malaria-endemic regions in Kenya.

ii. To derive the asymptotic properties of the proposed Bayesian GLL model.

iii. To evaluate the impact of maternal health and socio-economic factors on child survival in high-transmission malaria regions in Kenya.

1.4 Scope of the Study

This study contributes methodologically by developing and validating a Bayesian GLL survival model and empirically by applying it to real-world data on child mortality in malaria-endemic regions of Kenya. The research also focuses on understanding how maternal health behaviors, socioeconomic status, and regional malaria prevalence affect survival, with implications for public health planning and policy.

1.5 Significance of the Study

This study aims to apply advanced statistical modeling—the generalized log-logistic (GLL) model with Bayesian inference—to understand child survival in malaria-endemic regions of Kenya. By estimating how maternal health practices, socioeconomic factors, and malaria prevention efforts impact child mortality, the research offers valuable insights for designing targeted health interventions. This approach allows for more precise predictions of survival outcomes, supporting public health strategies aimed at reducing child mortality in high-risk areas.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews literature pertinent to child survival in malaria-endemic areas, focusing on the impact of maternal health factors, socioeconomic disparities, and regional differences on child mortality risks. The review also explores advanced survival models, particularly the generalized log-logistic (GLL) model, and Bayesian estimation methods to evaluate factors influencing child survival in Kenya. The literature underscores the challenge of high child mortality rates in malaria-prone regions due to malaria, socioeconomic inequalities, and health access disparities. Despite the existing survival studies, a gap remains in applying flexible survival models like GLL to assess non-monotonic hazard functions in child mortality in Kenya's malaria-endemic settings.

2.2 Malaria and Child Survival

Malaria continues to pose a severe health threat to children under five in sub-Saharan Africa, including Kenya, where high transmission rates increase mortality risks. Studies highlight that, in malaria-endemic regions, maternal health practices and malaria prevention measures significantly influence child survival outcomes. For instance, the World Health Organization (World Health Organization (2023)) indicates that malaria was responsible for approximately 260,000 child deaths annually, disproportionately impacting low-income families without access to preventive health measures such as bed nets or antimalarial treatments. Studies like (Amek et al. (2018)) have shown that malaria transmission intensity, measured by the entomological inoculation rate (EIR), is strongly associated with child mortality. (Bates

et al. (2021)) conducted a detailed scoping review on the prevalence and risk factors of malaria in children under five across sub-Saharan Africa, emphasizing the significance of socio-economic factors such as poverty and maternal education. Their research indicates that children in low-income households, particularly in rural areas, face increased malaria risk due to limited access to preventive resources like insecticide-treated bed nets (ITNs) and healthcare infrastructure. The study shows that these risk factors are key contributors to malaria-related mortality among children under five, highlighting the need for targeted interventions in socioeconomically disadvantaged regions. Similarly, (Mungai et al. (2019)) explored the role of maternal health practices on child survival rates in malaria-endemic areas of Kenya. They found that regular antenatal care visits, maternal knowledge of malaria prevention, and the use of bed nets were positively associated with improved survival outcomes for young children. However, the study notes a gap in modeling approaches that assess the long-term impact of these interventions on child mortality, especially when accounting for variables like maternal education and household income

2.3 Socioeconomic, Regional Disparities and Maternal Health Factors in Child Survival

Maternal health practices, socioeconomic status, and regional access disparities each play a critical role in child survival outcomes in malaria-prone regions of sub-Saharan Africa, including Kenya. Maternal health behaviors, such as attending antenatal care (ANC) visits and utilizing malaria prevention methods, are vital in reducing child mortality risks in these areas. Studies show that high frequencies of ANC visits are linked to lower mortality rates due to early detection and timely interventions for malaria during pregnancy (Omer and Smith (2020)). Furthermore, the use of insecticide-treated bed nets (ITNs) and prophylactic malaria treatments during pregnancy has demonstrated a statistically significant impact on child survival by lowering mortality rates (Gething and Noor (2019)). (Nyovani and Oriko (2020)) further emphasize that maternal education level also strongly influences these outcomes; mothers with higher educational attainment are more likely to seek preventative health care

and understand malaria prevention measures, which are essential to reducing child mortality in high-transmission areas.

In addition to maternal health factors, socioeconomic status and regional disparities exacerbate child mortality risks in malaria-endemic regions. Research by (Manda and Moindi (2021)) and (Mungai et al. (2019)) demonstrates that children from lower-income households face a heightened risk of malaria-related mortality due to limited access to healthcare services and preventive measures. Rural areas, in particular, experience significant health service accessibility gaps, contributing to higher mortality rates among children compared to urban areas. For instance, the Kenyan Demographic Health Survey (Kenya Demographic and Health Survey (2015)) highlights that infrastructure limitations in rural regions, such as inadequate healthcare facilities and limited malaria prevention resources, compound the effects of socioeconomic status and make timely health interventions challenging to deliver.

2.4 Survival Modeling Techniques in Child Mortality Studies

Survival analysis techniques are essential in examining factors influencing child mortality, especially in high-risk malaria regions. Traditional methods, particularly the Cox proportional hazards model, are widely utilized due to their interpretability and ability to assess hazard ratios. However, these models may struggle to account for the complex, non-linear risk patterns and time-dependent factors, such as seasonal malaria transmission and maternal health interventions, that are prevalent in malaria-endemic regions (Penfold and Zhang (2013)). To address these limitations, advanced survival models like the generalized log-logistic (GLL) model have been proposed, which offer the flexibility to capture both monotonic (constant or steadily increasing/decreasing) and non-monotonic (fluctuating) hazard functions. This is especially relevant in malaria studies, where child mortality risks do not follow a strictly linear or proportional pattern, often due to fluctuating malaria transmission rates and the episodic nature of malaria outbreaks.

Researchers such as (Zhang and Zhao (2019)) and (Langat and Ochieng (2022)) have highlighted the GLL model's advantages over traditional models in public health settings. In their study, (Zhang and Zhao (2019)) conducted a comparative study of survival models applied to tuberculosis datasets in China. Their analysis demonstrated that the GLL model provided a significantly better fit than the Cox model for datasets with non-linear hazards. Using maximum likelihood estimation methods, they found that the GLL model captured key variations in hazard rates influenced by treatment adherence and seasonal factors, leading to improved predictive accuracy and actionable insights for public health interventions. This demonstrated that the GLL model provides a superior fit for datasets characterized by varying hazard rates, as it accommodates non-linear trends that can emerge in survival data for complex diseases like malaria. (Langat and Ochieng (2022)) further showed that applying the GLL model in survival analysis allows for a more precise understanding of mortality dynamics in high-risk populations, particularly where hazard rates are influenced by multifactorial determinants.He applied the GLL model to study mortality in high-risk populations affected by HIV and opportunistic infections in sub-Saharan Africa. Through Bayesian estimation techniques, they incorporated prior knowledge about disease progression and socioeconomic disparities. Their findings underscored the GLL model's capacity to model survival outcomes in populations where hazards are influenced by multifactorial determinants such as poverty, geographic inaccessibility, and healthcare inequalities. This approach enabled the researchers to identify critical intervention points for reducing mortality in vulnerable groups. However, while these studies establish the efficacy of the GLL model, its application to child survival data in malaria-endemic areas, such as Kenya, remains limited. This study aims to address this gap by employing the GLL model to assess child mortality risk among children under five in malaria-prone Kenyan counties, providing a refined approach to survival analysis in such high-risk environments.

2.5 Bayesian Inference in Survival Analysis

Bayesian inference has increasingly gained prominence in survival analysis, particularly in public health research, as it allows researchers to incorporate prior knowledge into the model. This approach is particularly advantageous in settings with data sparsity or heterogeneous population characteristics, which is common in malaria-endemic regions where access to healthcare services and preventative measures varies widely (Gelman et al. (2014)). By including prior distributions based on past studies or expert opinions, Bayesian methods enhance the robustness of parameter estimates and improve predictive accuracy, making them well-suited for modeling survival in complex health datasets.

In their research, (Ahmed et al. (2021)) employed Bayesian inference to estimate parameters in a survival model aimed at understanding the mortality risk among children in resource-constrained settings. The study utilized Markov Chain Monte Carlo (MCMC) methods to calculate posterior distributions of survival parameters, ensuring a flexible integration of prior information on maternal health practices and socioeconomic disparities. Ahmed et al. reported that Bayesian models provided narrower credible intervals compared to frequentist confidence intervals, indicating more precise estimates. Their findings highlighted that incorporating prior knowledge about malaria prevention measures and antenatal care attendance significantly improved the model's ability to predict child survival outcomes in rural and under-resourced areas. (Kibuchi and Muthama (2020)) similarly focused on the application of Bayesian methods in survival analysis to model maternal and child health outcomes in sub-Saharan Africa. They applied hierarchical Bayesian models to account for regional variability and differences in healthcare access. The study demonstrated that Bayesian inference could effectively address uncertainties in observational health data by integrating regional covariates into the survival model. Kibuchi et al. found that socioeconomic factors, such as household income and education levels, were strong predictors of child mortality risks, with credible intervals revealing significant disparities across different geographic regions.

A key contribution in advanced survival modeling is the integration of spatial analysis using Bayesian frameworks, as detailed by (Egbon et al. (2022)). This approach utilizes Gaussian

Markov Random Fields (GMRFs) to model spatial heterogeneity, allowing researchers to account for regional variations in child survival outcomes. The spatial random effects are modeled with precision matrices that describe the dependency structure among geographical units, such as counties:

$$Q = \tau(D - W), \tag{2.1}$$

where:

i. τ represents the precision parameter,

ii. D is the diagonal degree matrix,

iii. W is the adjacency matrix indicating spatial relationships.

This framework effectively captures localized variations in child survival outcomes, which are critical in malaria-endemic regions.

To model survival probabilities, the paper by (Egbon et al. (2022)) employed a Bayesian survival function combining the likelihood for time-to-event data with prior distributions for parameters. The likelihood is expressed as:

$$L(y \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta)^{\delta_i} S(y_i \mid \theta)^{1-\delta_i}, \tag{2.2}$$

where:

i. $f(y_i \mid \theta)$ is the probability density function for event times,

ii. $S(y_i \mid \theta)$ is the survival function,

iii. δ_i is the event indicator (1 for event observed, 0 for censored),

iv. y is the observed survival data,

v. θ represents model parameters, including baseline hazards and covariate effects.

The Bayesian posterior distribution is then derived as:

$$\pi(\theta \mid y) \propto L(y \mid \theta) \cdot \pi(\theta), \tag{2.3}$$

where $\pi(\theta)$ represents the prior distribution for the parameters.

Another key contribution is from the work by (Muse et al. (2021)) significantly contributes to the literature on advanced survival modeling, particularly through its focus on the Bayesian and classical inference for the Generalized Log-Logistic (GLL) distribution. This distribution, as highlighted in the paper, extends the classical log-logistic distribution by incorporating an additional parameter to handle non-monotonic hazard functions, making it particularly suitable for survival data characterized by variability in hazard rates. The study outlines the distribution's key statistical properties, its application to real-world survival data, and its superior fit compared to traditional parametric models like Weibull and log-normal distributions.

Key Contributions

Mathematical Formulation

The hazard function for the GLL model is defined as:

$$h(t;\theta) = \frac{\alpha k(\eta t)^{\alpha-1}}{1+(\eta t)^{\alpha}}, \tag{2.4}$$

where:

i. α: shape parameter,

ii. k: scale parameter, and

iii. η: parameter controlling tail behavior.

This flexibility is vital for modeling survival in malaria-endemic regions where hazard rates are influenced by seasonal and socio-economic factors.

The survival function is:

$$S(t;\theta) = [1 + (\eta t)^{\alpha}]^{-k}. \tag{2.5}$$

Bayesian Estimation

(Muse et al. (2021)) applied Markov Chain Monte Carlo (MCMC) techniques to estimate the posterior distributions of the model parameters. The Bayesian framework enables the incorporation of prior knowledge, addressing data sparsity issues and enhancing predictive accuracy.

Model Validation

Using real-world data, the GLL distribution demonstrated better goodness-of-fit metrics compared to its submodels (e.g., Weibull, log-logistic) and other parametric distributions. This supports its application in public health for diseases like malaria.

Gap Identification

While the (Muse et al. (2021)) paper establishes the GLL distribution's versatility, it primarily focuses on general survival datasets without emphasizing specific malaria-related predictors such as antenatal care visits, ITN use, or regional socio-economic disparities. My research extends this framework by incorporating these domain-specific covariates into the Bayesian GLL model and examining its spatial and temporal dynamics in malaria-endemic Kenyan counties.

This study extends these findings by utilizing Bayesian inference for parameter estimation within the GLL model, ensuring that the model not only captures the unique mortality risks in malaria-endemic settings but also aligns with region-specific data from Kenyan counties. Specifically, this research seeks to estimate credible intervals and posterior distributions for key parameters, including maternal health factors, socioeconomic status, and malaria prevention measures. By applying Bayesian inference within the GLL framework, this study

aims to offer a novel approach to survival modeling in child mortality, addressing the need for more adaptive and regionally relevant methods in malaria-focused public health research.

Chapter 3

Methodology

3.1 Introduction

This section outlines the methodology employed to investigate child survival in malaria-endemic regions of Kenya using survival analysis techniques. The study utilizes child-level datasets from 2020 to model survival and mortality risks. Key methods include the generalized log-logistic (GLL) survival model and Bayesian inference, tailored to capture the complexities of child survival dynamics in high-risk areas.

3.2 Source of Data

This study draws upon publicly available data from Kenya's Demographic and Health Surveys (KDHS) conducted in 2020. This dataset provides comprehensive information on child health, maternal practices, socioeconomic factors, and malaria prevention measures. Data include variables such as child survival status, maternal health interventions, malaria prevention usage, and household characteristics. The study targets children under the age of five, with a focus on identifying mortality patterns and survival rates across malaria-endemic regions in Kenya. The Key data variables include:

i. **Outcome Variable**: Survival status (alive or deceased) and time-to-event (measured in months since birth).

ii. **Explanatory Variables**: Maternal education, household wealth index, malaria prevention measures (e.g., ITNs, antimalarial use), regional healthcare access, and antenatal care visits.

3.3 Data Preprocessing

3.3.1 Handling Missing Values

Missing data is a common challenge in large-scale health surveys, and the following approaches will be employed to address it: First, exploratory analysis will identify the extent and patterns of missingness across variables, such as maternal health practices and socio-economic indicators, and determine whether the data are Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR). For low levels of missingness , mean or mode imputation will be applied for continuous and categorical variables, respectively, while moderate to high levels of missingness will be handled using multiple imputation techniques, such as predictive mean matching or Bayesian imputation, to ensure unbiased estimates. Finally, sensitivity analysis will be conducted to compare results from analyses with and without imputed data, assessing the robustness of the findings.

3.4 Model Description

We employ the Generalized Log-Logistic (GLL) model to analyze child survival dynamics. The GLL model is well-suited for capturing non-linear and time-varying hazard functions, which are prevalent in malaria-endemic regions. Unlike conventional Cox models, the GLL model can capture both monotonic and non-monotonic hazard functions, making it ideal for diseases like malaria, where mortality risks fluctuate due to seasonal transmission patterns and healthcare access variability.

The hazard function of the GLL model is given by:

$$h(t) = \frac{\gamma \cdot \lambda^{\gamma} \cdot t^{\gamma-1}}{1 + (\lambda \cdot t)^{\gamma}} \tag{3.1}$$

where:

i. λ: Scale parameter, representing baseline survival risk.

ii. γ: Shape parameter, controlling the non-linearity and risk trend.

iii. t: Time variable, representing the child's age in months.

3.4.1 Survival Function

The survival function $S(t)$ for the GLL model is:

$$S(t;\lambda,\gamma) = \frac{1}{1+(\lambda t)^{\gamma}}, \tag{3.2}$$

where $\lambda > 0$ and $\gamma > 0$ are scale and shape parameters, respectively.

3.4.2 Likelihood Function

For a dataset with n observations, where t_i represents observed time and δ_i is the event indicator ($\delta_i = 1$ for death, $\delta_i = 0$ for censoring), the likelihood function is:

$$L(\lambda,\gamma) = \prod_{i=1}^{n} \left[h(t_i;\lambda,\gamma)^{\delta_i} \cdot S(t_i;\lambda,\gamma)^{1-\delta_i} \right]. \tag{3.3}$$

Taking the natural logarithm, the log-likelihood function is:

$$\log L(\lambda,\gamma) = \sum_{i=1}^{n} \left[\delta_i \log h(t_i;\lambda,\gamma) + (1-\delta_i) \log S(t_i;\lambda,\gamma) \right]. \tag{3.4}$$

This represents the total likelihood function used for parameter estimation.

In this study, the GLL model will be adapted to evaluate survival outcomes by: Incorporating covariates such as maternal health practices, malaria prevention measures, and socio-economic factors directly into the hazard function to refine risk estimation.

This approach extends previous applications of the GLL model Zhang and Zhao (2019) by explicitly focusing on child mortality data in malaria-endemic regions. This study will estimate the model parameters using Bayesian inference techniques, combining prior distributions with observed data to evaluate the influence of maternal health practices, malaria

prevention measures, and socio-economic factors on child survival. The analysis will focus on the 2020 dataset to provide detailed insights into survival dynamics in malaria-endemic regions of Kenya.

3.5 Bayesian Inference for the GLL Model

In this study, **Bayesian hierarchical modeling** is used to estimate the parameters of the GLL model while accounting for both individual-level (child-specific) and regional-level (area-specific) influences on child survival. The hierarchical model involves two levels:

i. **Level 1**: Individual-level data, such as maternal health, malaria prevention, and socio-economic factors.

ii. **Level 2**: Regional-level data, such as regional malaria prevalence and socio-economic conditions.

Bayesian Framework for Parameter Estimation The **Bayesian inference** process combines prior knowledge (through **priors**) with observed data to estimate the posterior distributions of model parameters. The posterior distribution incorporates all available information and reflects updated beliefs about the parameters given the data.

The likelihood function for the Bayesian GLL model is:

$$L(\theta \mid Y) = \prod_{i=1}^{n} \left[h(t_i \mid \theta)^{\delta_i} \cdot S(t_i \mid \theta)^{1-\delta_i} \right], \tag{3.5}$$

where:

i. δ_i: event indicator (1 = death, 0 = censored),

ii. $S(t)$: survival function,

iii. θ: vector of parameters (λ, γ, β).

The prior distribution for the parameters γ and λ reflects prior beliefs about their values before observing the data. The prior distributions are specified as follows:

a. **Prior for γ:** A Gamma prior is chosen to reflect positive values for γ, which controls the shape of the hazard function:

$$\gamma \sim \text{Gamma}(\alpha, \beta)$$

where α and β are hyperparameters chosen based on prior knowledge.

b. **Prior for λ:** A log-normal or exponential prior can be used for λ, the scale parameter, which is also positive:

$$\lambda \sim \text{Log-Normal}(\mu, \sigma)$$

or

$$\lambda \sim \text{Exponential}(\lambda_0)$$

where μ, σ, and λ_0 are hyperparameters that can be set based on prior knowledge or data.

Once the priors are specified, the posterior distribution for γ and λ is obtained using Bayes' Theorem:

$$p(\gamma, \lambda \mid t, \delta) \propto L(\gamma, \lambda \mid t, \delta) \cdot p(\gamma) \cdot p(\lambda)$$

where $p(\gamma)$ and $p(\lambda)$ are the priors, and $L(\gamma, \lambda \mid t, \delta)$ is the likelihood function.

Hierarchical Structure

Incorporating hierarchical modeling involves introducing random effects for regional-level parameters. This allows the model to account for variation in survival patterns across different regions of Kenya, based on factors such as regional malaria prevalence, healthcare infrastructure, and socio-economic conditions. The model can be specified as follows:

Child-level Model (Level 1) Each child's hazard function is:

$$h(t_i) = \gamma \cdot \lambda^{\gamma} \cdot t_i^{\gamma-1} \left(1 + (\lambda \cdot t_i)^{\gamma}\right). \tag{3.6}$$

This formulation links the hazard function directly to the likelihood.

Regional-level Model (Level 2)

Regional variations in survival are captured by modeling parameters as random effects:

$$\lambda_r \sim \text{Normal}(\mu_\lambda, \sigma_\lambda), \tag{3.7}$$

$$\gamma_r \sim \text{Normal}(\mu_\gamma, \sigma_\gamma). \tag{3.8}$$

This hierarchical structure accounts for regional disparities in survival dynamics.

For child i in region r, the hazard function is:

$$h_r(t_{ij}) = \frac{\gamma_r \lambda_r^{\gamma_r} t_{ij}^{\gamma_r - 1}}{1 + (\lambda_r t_{ij})^{\gamma_r}} \tag{3.9}$$

where:

i. λ_r: Region-specific scale parameter (baseline risk)

ii. γ_r: Region-specific shape parameter (risk trend)

iii. t_{ij}: Survival time for child i in region j

The population-Averaged survival is given by:

$$S_r(t_{ij}|\lambda_r, \gamma_r) = \frac{1}{1 + (\lambda_r t_{ij})^{\gamma_r}} \tag{3.10}$$

Marginal Likelihood

The integrated likelihood across all regions is:

$$L(\theta) = \prod_{r=1}^{R} \int \int \prod_{i=1}^{n_r} L_{ij}(\lambda_r, \gamma_r) f(\lambda_r | \mu_\lambda, \sigma_\lambda^2) f(\gamma_r | \mu_\gamma, \sigma_\gamma^2) d\lambda_r d\gamma_r \tag{3.11}$$

where:

i. $L_{ij}(\lambda_r, \gamma_r) = [h(t_{ij}|\lambda_r, \gamma_r)]^{\delta_{ij}} \cdot [S(t_{ij}|\lambda_r, \gamma_r)]^{1-\delta_{ij}}$

ii. δ_{ij} is the event indicator (1 if death, 0 if censored)

Population-Averaged Survival

$$S_{\text{avg}}(t) = \frac{1}{R} \sum_{r=1}^{R} S_r(t) \tag{3.12}$$

Implementation and Inference

To estimate the parameters and assess their posterior distributions, we utilized the R programming language in combination with Stan, a state-of-the-art probabilistic programming framework for Bayesian inference. Stan employs Hamiltonian Monte Carlo (HMC) and its variant, No-U-Turn Sampler (NUTS), to efficiently explore the posterior distributions of the model parameters. This approach provides robust and efficient sampling, allowing for the estimation of key parameters, including individual-level effects and regional-level random effects.

3.5.1 Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS)

3.5.2 Hamiltonian Monte Carlo (HMC)

Hamiltonian Monte Carlo (HMC) is an advanced Markov Chain Monte Carlo (MCMC) method designed to improve sampling efficiency by leveraging concepts from classical physics. Unlike traditional MCMC algorithms such as the Metropolis-Hastings algorithm, which suffer from random walk behavior, HMC uses auxiliary momentum variables to guide sampling.

HMC is based on Hamiltonian dynamics, where the system's total energy is defined as:

$$H(\theta, p) = U(\theta) + K(p), \tag{3.13}$$

where:

i. θ represents the parameters of interest.

ii. p represents the auxiliary momentum variables.

iii. $U(\theta) = -\log p(\theta|\text{Data})$ is the potential energy, derived from the log posterior.

iv. $K(p) = \frac{1}{2} p^T M^{-1} p$ is the kinetic energy, typically modeled as a multivariate Gaussian distribution with mass matrix M.

To sample efficiently from the posterior distribution, HMC solves Hamilton's equations:

$$\frac{d\theta}{dt} = \frac{\partial K}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial U}{\partial \theta}. \tag{3.14}$$

The key steps in HMC include:

i. **Initialization:** Set initial values $\theta^{(0)}$ and draw initial momentum $p \sim \mathcal{N}(0, M)$.

ii. **Leapfrog Integration:** Approximate the continuous evolution of the system using small discrete steps of size ε.

iii. **Metropolis Acceptance Step:** Accept or reject the proposed state based on energy conservation.

HMC allows for efficient exploration of high-dimensional posterior distributions by making long-distance proposals while maintaining a high acceptance rate.

3.5.3 No-U-Turn Sampler (NUTS)

While HMC improves sampling efficiency, it requires the user to specify the number of leapfrog steps, which can significantly affect performance. The No-U-Turn Sampler (NUTS) is an extension of HMC that automatically selects the optimal number of leapfrog steps to ensure efficient posterior exploration.

NUTS adaptively determines the trajectory length by monitoring the movement of the simulated particle in parameter space. The sampler stops the trajectory when it detects a U-turn, meaning the sampler would start retracing its steps. This prevents unnecessary computation and improves efficiency.

Key advantages of NUTS:

i. **Automatic Tuning:** Eliminates the need for manual selection of the leapfrog step count.

ii. **Efficient Exploration:** Avoids inefficient random walks and improves posterior sampling.

iii. **Reduced Autocorrelation:** Produces highly effective samples with low correlation, improving convergence diagnostics.

NUTS ensures efficient and adaptive exploration of high-dimensional posterior distributions, making it particularly useful for hierarchical Bayesian models.

3.5.4 Implementation in Stan

Stan's implementation of NUTS automatically adapts the step size ε and the mass matrix M during warm-up to optimize sampling efficiency. The user specifies a model in Stan's probabilistic programming language, and Stan generates efficient HMC/NUTS-based inference.

The Bayesian hierarchical survival model was implemented using the **rstan** package in R, which provides an interface to Stan for model specification, fitting, and posterior sampling. Prior distributions were specified based on domain knowledge, ensuring a well-regularized estimation process. Posterior summaries—including means, standard deviations, and credible intervals—were obtained to interpret the effects of covariates on child survival.

To assess the convergence of the MCMC chains, we examined **trace plots**, **R-hat values**, and **effective sample sizes (ESS)**. Additionally, **posterior predictive checks (PPCs)** were conducted to validate the model's fit to the observed data. The combination of **Stan's HMC/NUTS sampling** and **rigorous diagnostics** ensures the reliability of the posterior estimates and enhances the interpretability of the Bayesian Generalized Log-Logistic (GLL) model.

Application of the Model to the Data

The following variables from the dataset were used in the analysis:

Child-level variables:

1. Maternal health: Number of antenatal care visits, malaria prevention during pregnancy (e.g., SP/Fansidar), education level.

2. Household malaria prevention: Usage of insecticide-treated nets, whether children slept under a mosquito net the previous night.

3. Socio-economic factors: Wealth index.

4. Time-to-event: Age of the child in months (time to death or censoring).

5. Censoring indicator: Whether the child is alive or dead.

Regional-level variables:

1. Malaria prevalence: Regional malaria prevalence rates.

2. Socio-economic conditions: Regional wealth index,

These variables are incorporated into the hierarchical Bayesian model, with the child-level covariates influencing the individual survival times and the regional-level covariates accounting for spatial variations in survival outcomes.

Derivation of Asymptotic Properties for Model Performance

To assess the theoretical performance of the proposed **Bayesian Generalized Log-Logistic (GLL) survival model**, we derive its **asymptotic properties**, including bias, variance, and mean squared error (MSE). These derivations provide insight into the long-run behavior of the estimators as the sample size increases. In particular, we investigate key properties of the Bayesian estimators, such as:

i. **Unbiasedness:** Whether the estimator's expected value equals the true parameter value.

ii. **Consistency:** Whether the estimator converges in probability to the true parameter as the sample size grows.

iii. **Efficiency:** Whether the estimator achieves the lowest possible variance (i.e., the Cramér-Rao lower bound) in the asymptotic limit.

iv. **Sufficiency:** Whether the estimator utilizes all the information available in the data regarding the parameter.

Establishing these properties ensures that our model is robust and reliable for practical applications in analyzing child survival data.

Asymptotic Analysis Framework

The derivation process follows these steps:

Parameter Estimation in Large-Sample Theory

i. Using Bayesian estimation principles, we analyze the asymptotic behavior of the posterior distributions of λ, γ, and covariates (β).

ii. By applying Laplace's approximation and invoking asymptotic normality, we assess how these estimates converge as the sample size grows.

Bias and Variance Derivations

i. The bias of the Bayesian estimator $\hat{\theta}$ (where θ represents a parameter such as λ or γ) is defined as:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta, \tag{3.15}$$

where $E[\hat{\theta}]$ is the posterior mean and θ is the true parameter value.

ii. The asymptotic variance is derived using the Fisher information matrix. This variance reflects the estimator's efficiency in large samples.

Mean Squared Error (MSE) Analysis

i. The overall accuracy of the estimator is quantified by the Mean Squared Error (MSE), defined as:

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}^2(\hat{\theta}). \tag{3.16}$$

ii. For large sample sizes, the bias term becomes negligible, and the MSE is dominated by the variance component.

The posterior distribution of the parameters $\theta = \{\theta_r\}_{r=1}^{R}$ and hyperparameters $\phi = (\mu_\lambda, \mu_\gamma, \sigma_\lambda^2, \sigma_\gamma^2)$ is:

$$p(\theta, \phi \mid \mathscr{D}) \propto L(\theta \mid \mathscr{D}) \cdot p(\theta \mid \phi) \cdot p(\phi), \tag{3.17}$$

where:

i. $L(\theta \mid \mathscr{D})$ is the likelihood function,

ii. $p(\theta \mid \phi)$ is the prior distribution of θ given ϕ,

iii. $p(\phi)$ is the hyperprior distribution.

The likelihood function is:

$$L(\theta \mid \mathscr{D}) = \prod_{r=1}^{R} \prod_{i=1}^{n_r} \left[h(t_i \mid \theta_r)^{\delta_i} \cdot S(t_i \mid \theta_r)^{1-\delta_i} \right], \tag{3.18}$$

where n_r is the number of children in region r.

—

3.5.5 Posterior Mode Estimation

The posterior mode $(\hat{\theta}, \hat{\phi})$ is obtained by maximizing the log-posterior:

$$(\hat{\theta}, \hat{\phi}) = \arg\max_{\theta, \phi} \ell(\theta, \phi), \tag{3.19}$$

where $\ell(\theta, \phi) = \log p(\theta, \phi \mid \mathscr{D})$ is the log-posterior.

The log-posterior is given by:

$$\ell(\theta, \phi) = \log() + \log(p(\theta \mid \phi)) + \log(p(\phi)). \tag{3.20}$$

To find the posterior mode, we solve the following system of equations:

$$\frac{\partial \ell(\theta, \phi)}{\partial \theta} = 0, \quad \frac{\partial \ell(\theta, \phi)}{\partial \phi} = 0. \tag{3.21}$$

Derivatives of the Log-Posterior

Derivative of the Log-Likelihood: The log-likelihood is:

$$\log L(\theta \mid \mathscr{D}) = \sum_{r=1}^{R} \sum_{i=1}^{n_r} \left[\delta_i \log h(t_i \mid \theta_r) + (1 - \delta_i) \log S(t_i \mid \theta_r) \right]. \tag{3.22}$$

The derivative with respect to $\theta_r = (\lambda_r, \gamma_r)$ is:

$$\frac{\partial \log L(\theta \mid \mathscr{D})}{\partial \theta_r} = \sum_{i=1}^{n_r} \left[\delta_i \frac{\partial \log h(t_i \mid \theta_r)}{\partial \theta_r} + (1 - \delta_i) \frac{\partial \log S(t_i \mid \theta_r)}{\partial \theta_r} \right]. \tag{3.23}$$

For the GLL model, the hazard function $h(t_i \mid \theta_r)$ and survival function $S(t_i \mid \theta_r)$ are:

$$h(t_i \mid \theta_r) = \frac{\gamma_r \lambda_r^{\gamma_r} t_i^{\gamma_r - 1}}{1 + (\lambda_r t_i)^{\gamma_r}}, \tag{3.24}$$

$$S(t_i \mid \theta_r) = \frac{1}{1 + (\lambda_r t_i)^{\gamma_r}}. \tag{3.25}$$

The derivatives of $\log h(t_i \mid \theta_r)$ and $\log S(t_i \mid \theta_r)$ with respect to λ_r and γ_r can be computed using the chain rule.

Derivative of the Log-Prior: The log-prior for θ_r is:

$$\log p(\theta_r \mid \phi) = \log p(\lambda_r \mid \mu_\lambda, \sigma_\lambda^2) + \log p(\gamma_r \mid \mu_\gamma, \sigma_\gamma^2). \tag{3.26}$$

For the log-normal priors:

$$\log(p(\lambda_r \mid \mu_\lambda, \sigma_\lambda^2)) = -\frac{(\log \lambda_r - \mu_\lambda)^2}{2\sigma_\lambda^2} - \log \lambda_r - \frac{1}{2} \log(2\pi\sigma_\lambda^2), \tag{3.27}$$

$$\log(p(\gamma_r \mid \mu_\gamma, \sigma_\gamma^2)) = -\frac{(\log \gamma_r - \mu_\gamma)^2}{2\sigma_\gamma^2} - \log \gamma_r - \frac{1}{2} \log(2\pi\sigma_\gamma^2). \tag{3.28}$$

The derivatives with respect to λ_r and γ_r are:

$$\frac{\partial \log p(\lambda_r \mid \mu_\lambda, \sigma_\lambda^2)}{\partial \lambda_r} = -\frac{\log \lambda_r - \mu_\lambda}{\lambda_r \sigma_\lambda^2} - \frac{1}{\lambda_r}, \tag{3.29}$$

$$\frac{\partial \log p(\gamma_r \mid \mu_\gamma, \sigma_\gamma^2)}{\partial \gamma_r} = -\frac{\log \gamma_r - \mu_\gamma}{\gamma_r \sigma_\gamma^2} - \frac{1}{\gamma_r}. \tag{3.30}$$

Derivative of the Log-Hyperprior: The log-hyperprior for $\phi = (\mu_\lambda, \mu_\gamma, \sigma_\lambda^2, \sigma_\gamma^2)$ is:

$$\log(p(\phi)) = \log p(\mu_\lambda) + \log p(\mu_\gamma) + \log p(\sigma_\lambda^2) + \log p(\sigma_\gamma^2). \tag{3.31}$$

For the normal and inverse-gamma hyperpriors:

$$\log(p(\mu_\lambda)) = -\frac{(\mu_\lambda - m_\lambda)^2}{2s_\lambda^2} - \frac{1}{2}\log(2\pi s_\lambda^2), \tag{3.32}$$

$$\log(p(\sigma_\lambda^2)) = -a_\lambda \log \sigma_\lambda^2 - \frac{b_\lambda}{\sigma_\lambda^2} - \log\Gamma(a_\lambda) + a_\lambda \log b_\lambda. \tag{3.33}$$

The derivatives with respect to μ_λ and σ_λ^2 are:

$$\frac{\partial \log p(\mu_\lambda)}{\partial \mu_\lambda} = -\frac{\mu_\lambda - m_\lambda}{s_\lambda^2}, \tag{3.34}$$

$$\frac{\partial \log p(\sigma_\lambda^2)}{\partial \sigma_\lambda^2} = -\frac{a_\lambda}{\sigma_\lambda^2} + \frac{b_\lambda}{(\sigma_\lambda^2)^2}. \tag{3.35}$$

Numerical Optimization

The system of equations:

$$\frac{\partial \ell(\theta, \phi)}{\partial \theta} = 0, \quad \frac{\partial \ell(\theta, \phi)}{\partial \phi} = 0, \tag{3.36}$$

is typically solved using numerical methods such as:

Newton-Raphson Method: - Iteratively update the parameters using:

$$(\theta^{(k+1)}, \phi^{(k+1)}) = (\theta^{(k)}, \phi^{(k)}) - \left[\ell''(\theta^{(k)}, \phi^{(k)})\right]^{-1} \nabla \ell(\theta^{(k)}, \phi^{(k)}), \tag{3.37}$$

where $\ell''(\theta, \phi)$ is the Hessian matrix of second derivatives.

Laplace Approximation

To approximate the posterior distribution, we perform a **second-order Taylor expansion** of the log-posterior $\ell(\theta, \phi)$ around the posterior mode $(\hat{\theta}, \hat{\phi})$:

$$\ell(\theta, \phi) \approx \ell(\hat{\theta}, \hat{\phi}) + \frac{1}{2}(\theta - \hat{\theta}, \phi - \hat{\phi})^T \ell''(\hat{\theta}, \hat{\phi})(\theta - \hat{\theta}, \phi - \hat{\phi}), \tag{3.38}$$

where $\ell''(\hat{\theta}, \hat{\phi})$ is the Hessian matrix of second derivatives evaluated at $(\hat{\theta}, \hat{\phi})$.

Exponentiating this approximation, the posterior density becomes:

$$p(\theta, \phi \mid \mathscr{D}) \approx \mathscr{N}\left((\hat{\theta}, \hat{\phi}), -\ell''(\hat{\theta}, \hat{\phi})^{-1}\right). \tag{3.39}$$

Bias Expansion

The bias of the posterior mode $(\hat{\theta}, \hat{\phi})$ is derived using a **higher-order Taylor expansion** around the true parameter values (θ_0, ϕ_0):

$$(\hat{\theta} - \theta_0, \hat{\phi} - \phi_0) \approx -\ell''(\theta_0, \phi_0)^{-1} \nabla \ell(\theta_0, \phi_0) - \frac{1}{2}\ell''(\theta_0, \phi_0)^{-1} \ell'''(\theta_0, \phi_0)(\hat{\theta} - \theta_0, \hat{\phi} - \phi_0)^2. \tag{3.40}$$

Taking expectations and using the fact that $E[\nabla \ell(\theta_0, \phi_0)] = 0$, the bias is:

$$E[(\hat{\theta} - \theta_0, \hat{\phi} - \phi_0)] \approx -\frac{1}{2}\ell''(\theta_0, \phi_0)^{-1} E\left[\ell'''(\theta_0, \phi_0)(\hat{\theta} - \theta_0, \hat{\phi} - \phi_0)^2\right] \tag{3.41}$$

For large n, the bias is of order $O(n^{-1})$:

$$\text{Bias}(\hat{\theta}, \hat{\phi}) = O(n^{-1}). \tag{3.42}$$

Posterior Mean $\bar{\theta}$:

$$E[\bar{\theta} - \theta_0] = \frac{1}{n} I^{-1}(\theta_0) \nabla b(\theta_0) + O(n^{-2})$$

Where: $b(\theta_0)$ is the prior's effect (vanishes for flat priors).

Thus, the estimator is asymptotically unbiased.

Efficiency : Variance via Fisher Information

The Fisher information matrix $I(\theta_0, \phi_0)$ is defined as:

$$I(\theta_0, \phi_0) = -E\left[\frac{\partial^2 \log L(\theta, \phi \mid \mathscr{D})}{\partial(\theta, \phi)\partial(\theta, \phi)^T}\right]. \tag{3.43}$$

From the Laplace approximation, the Hessian of the log-posterior is:

$$\ell''(\hat{\theta}, \hat{\phi}) \approx -nI(\theta_0, \phi_0). \tag{3.44}$$

Thus, the posterior covariance matrix is:

$$\mathrm{Var}(\hat{\theta}, \hat{\phi}) \approx -\ell''(\hat{\theta}, \hat{\phi})^{-1} = \frac{1}{n} I^{-1}(\theta_0, \phi_0). \tag{3.45}$$

By the Cramér-Rao Lower Bound, any unbiased estimator has variance bounded by:

$$\mathrm{Var}(\hat{\theta}) \geq I^{-1}(\theta_0, \phi_0). \tag{3.46}$$

Since the variance of $\hat{\theta}$ attains this bound for large n, the estimator is efficient. —

Sufficiency

Theorem 3.1 (Asymptotic Sufficiency via Bernstein-von Mises)**.**

Under the regularity conditions:

1. Differentiability in quadratic mean at θ_0,
2. Local Asymptotic Normality (LAN) of the log-likelihood,
3. Prior continuity $p(\theta) > 0$ in a neighborhood of θ_0,

the posterior distribution satisfies:

$$\sup_{A} \left| \Pi\left(\sqrt{n}(\theta - \hat{\theta}_{\text{MLE}}) \in A \mid \mathscr{D}\right) - N(0, I^{-1}(\theta_0))(A) \right| \xrightarrow{P} 0, \tag{3.47}$$

where:

i. $\hat{\theta}_{\text{MLE}}$ is the maximum likelihood estimator,

ii. $I(\theta_0)$ is the Fisher information matrix.

Theorem 3.2 (Bayesian Asymptotic Sufficiency)**.** *Under the regularity conditions of Theorem 3.1, the posterior distribution* $\Pi(\theta \mid \mathscr{D})$ *isasymptotically sufficient for* θ

in the sense that:

1. For any bounded continuous function $f(\theta)$,

$$\mathbb{E}[f(\theta) \mid \mathscr{D}] = \mathbb{E}[f(\theta) \mid \hat{\theta}_{\text{MLE}}] + o_{\mathbb{P}}(1) \tag{3.48}$$

2. The Fisher information bound is achieved:

$$\text{Var}(\theta \mid \mathscr{D}) = [nI(\hat{\theta}_{\text{MLE}})]^{-1} + o_{\mathbb{P}}(n^{-1}) \tag{3.49}$$

3. The data $\mathscr{D}$ influences inference only through $\hat{\theta}_{\text{MLE}}$ in the limit:

$$\lim_{n\to\infty} D_{\text{KL}}\left(\Pi(\theta \mid \mathscr{D}) \middle\| \Pi(\theta \mid \hat{\theta}_{\text{MLE}})\right) = 0 \tag{3.50}$$

Proof. The result follows from:

1. The Bernstein–von Mises theorem's normal approximation.

2. The asymptotic equivalence of posterior moments and MLE-based estimates.

3. Pinsker's inequality controlling the Kullback–Leibler divergence.

□

Interpretation Thus, the posterior distribution $\Pi(\theta \mid \mathscr{D})$ is asymptotically sufficient for θ because:

i. It concentrates around $\hat{\theta}_{\text{MLE}}$ at a $\sqrt{n}$-rate.

ii. All Bayesian inferences (point estimates, credible sets) depend on $\mathscr{D}$ only through $\hat{\theta}_{\text{MLE}}$ for large n.

iii. No additional information about θ is contained in $\mathscr{D}$ beyond what $\hat{\theta}_{\text{MLE}}$ provides.

This establishes that while finite-sample exact sufficiency may not hold for GLL models, Bayesian inference naturally recovers sufficiency in the asymptotic regime through posterior concentration.

Mean Squared Error (MSE) Analysis

The mean squared error (MSE) of the posterior mode $(\hat{\theta}, \hat{\phi})$ is:

$$\text{MSE}(\hat{\theta}, \hat{\phi}) = E\left[(\hat{\theta} - \theta_0, \hat{\phi} - \phi_0)^2\right] = \text{Var}(\hat{\theta}, \hat{\phi}) + \left[\text{Bias}(\hat{\theta}, \hat{\phi})\right]^2. \tag{3.51}$$

Substituting the variance and bias results:

$$\text{MSE}(\hat{\theta}, \hat{\phi}) \approx \frac{1}{n} I^{-1}(\theta_0, \phi_0) + O(n^{-2}). \tag{3.52}$$

For large n, the bias term becomes negligible, and the MSE is dominated by the variance term:

$$\text{MSE}(\hat{\theta}, \hat{\phi}) \approx \frac{1}{n} I^{-1}(\theta_0, \phi_0). \tag{3.53}$$

—

Consistency : convergence in Probability

Under regularity conditions, the posterior distribution $p(\theta \mid \mathscr{D})$ concentrates around the true parameter θ_0 as $n \to \infty$: By the definition of consistency, we need to show:

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \quad \text{as } n \to \infty.$$

Given that

$$\text{Var}(\hat{\theta}_n) = O(n^{-1}),$$

we have:

$$\Pr(|\hat{\theta}_n - \theta_0| > \varepsilon) \leq \frac{\text{Var}(\hat{\theta}_n)}{\varepsilon^2} = O(n^{-1}).$$

Since $O(n^{-1}) \to 0$ as $n \to \infty$, we conclude:

$$\hat{\theta}_n \xrightarrow{P} \theta_0.$$

is consistent.

The explicit derivations confirm that:

1. Asymptotic Normality: The posterior distribution is asymptotically normal:

$$p(\theta, \phi \mid \mathscr{D}) \approx \mathscr{N}\left((\hat{\theta}, \hat{\phi}), \frac{1}{n} I^{-1}(\theta_0, \phi_0)\right). \tag{3.54}$$

   This ensures that Bayesian estimators exhibit the same asymptotic properties as classical maximum likelihood estimators.

2. Bias: The bias of the posterior mode is of order $O(n^{-1})$, implying that for large n, the estimator is approximately unbiased:

$$E[(\hat{\theta} - \theta_0, \hat{\phi} - \phi_0)] = O(n^{-1}).$$

3. Consistency: The estimator satisfies:

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \quad \text{as } n \to \infty.$$

   This follows from the fact that the posterior variance shrinks at the rate $O(n^{-1})$, ensuring that the estimator converges in probability to the true parameter.

4. Variance: The variance of the posterior mode is of order $O(n^{-1})$, aligning with efficiency bounds:
$$\text{Var}(\hat{\theta}, \hat{\phi}) \approx \frac{1}{n} I^{-1}(\theta_0, \phi_0).$$

5. Mean Squared Error (MSE): The MSE of the posterior mode is dominated by the variance term for large n:
$$\text{MSE}(\hat{\theta}, \hat{\phi}) \approx \frac{1}{n} I^{-1}(\theta_0, \phi_0). \tag{3.55}$$

6. Sufficiency: The Bernstein-von Mises theorem establishes that the posterior distribution is asymptotically sufficient.

7. Efficiency: The Bayesian estimator attains the asymptotic Cramér-Rao lower bound, ensuring minimal variance among unbiased estimators:
$$\text{Var}(\hat{\theta}) \geq [I(\theta)]^{-1}.$$

   The posterior covariance matrix follows the inverse Fisher information, supporting the estimator's efficiency in large samples.

These results hold under the hierarchical Bayesian framework, where regional-level random effects are incorporated into the model. The Bayesian GLL model provides a robust and flexible approach for estimating survival parameters while accounting for complex hierarchical dependencies in the data. The consistency, efficiency, and sufficiency of the estimator confirm its reliability for modeling child survival in malaria-endemic regions.

Evaluation of the Model

After fitting the model using MCMC sampling in Rstan, the posterior distributions of the parameters were thoroughly examined to assess the uncertainty and reliability of the parameter estimates. The credible intervals for each parameter offered insights into the

range of plausible values, integrating both prior knowledge and observed data. The model's goodness-of-fit was evaluated using posterior predictive checks, comparing simulated data generated from the model to the observed survival data. This approach ensured that the model captures the underlying patterns in the data accurately.

3.5.6 Posterior Predictive Validation

To complement theoretical derivations, we performed posterior predictive checks by comparing the observed data distribution with the model's predicted survival times. This ensures that the theoretical asymptotic properties align with empirical performance.

Conclusions

This methodology provides a comprehensive approach to understanding child survival in malaria-endemic regions of Kenya, using the Generalized Log-Logistic (GLL) model with Bayesian hierarchical inference. By accounting for both individual-level and regional-level factors, this model offers flexibility in capturing complex, non-monotonic hazard patterns over time. The Bayesian framework allows for the integration of prior knowledge and uncertainty, leading to more accurate and robust estimates of survival probabilities and mortality risk.

3.5.7 Ethical Considerations

For the purpose of this study, permission by the SU-ISERC was obtained after the proposal defence was approved. Clearance was obtained for the project as well as developments from it.

Chapter 4

Results and Interpretation

4.1 Introduction

This chapter presents the results of the exploratory data analysis, theoretical foundations and asymptotic properties, diagnostic tests, and findings from the fitted Bayesian Generalized Log logistic model to assess the impact of maternal health and socio-economic factors on child survival in high-transmission malaria regions in Kenya.

4.2 Frequency distribution

Table 4.1 presents the frequency distribution of key socio-demographic and health-related variables in the study. The data covers maternal education levels, child survival status, wealth index, malaria prevention practices, regional distribution, malaria endemicity, reported malaria cases, and place of residence. These distributions provide insights into the characteristics of the study population, highlighting potential disparities in education, wealth, healthcare access, and malaria exposure. The findings serve as a foundation for further analysis on child health outcomes and associated risk factors.

Variable	Category	Count	Percentage (%)
Highest Educational Level	No education	546	15.06
	Primary	1591	43.88
	Secondary	1062	29.29
	Higher	427	11.78
Child is Alive	No	45	1.24
	Yes	3581	98.76
Wealth Index Combined	Poorest	654	18.04
	Poorer	730	20.13
	Middle	1032	28.46
	Richer	726	20.02
	Richest	484	13.35
Took SP/Fansidar for Malaria During Pregnancy	No	1528	42.14
	Yes	2098	57.86
Region	Coast	472	13.02
	North Eastern	162	4.47
	Eastern	484	13.35
	Central	201	5.54
	Rift Valley	835	23.03
	Western	681	18.78
	Nyanza	754	20.79
	Nairobi	37	1.02
Malaria Endemicity Zone	Highland Epidemic Prone	809	22.31
	Lake Endemic	1364	37.62
	Coastal Endemic	413	11.39
	Seasonal	1040	28.68
Told Child Had Malaria	No	2664	73.47
	Yes	962	26.53
Type of Place of Residence	Urban	1371	37.81
	Rural	2255	62.19

Table 4.1: Frequency Distribution of Key Variables

Educational Level: Most mothers have primary education (43.88%), followed by secondary education (29.29%). A notable 15.06% lack formal education, which may impact their healthcare decisions. Only 11.78% have higher education, potentially contributing to better maternal health awareness.

Child Survival: The dataset indicates a low child mortality rate, with 98.76% of children alive and 1.24% deceased.

Wealth Index: A significant portion of households belong to the middle class (28.46%), while 20.13% fall into the poorer category. The poorest group makes up 18.04%, highlighting economic disparities. Only 13.35% are in the wealthiest segment.

Use of SP/Fansidar for Malaria Prevention During Pregnancy: While 57.86% of mothers took SP/Fansidar to prevent malaria, 42.14% did not, possibly due to lack of awareness or access.

Regional Distribution: Rift Valley has the highest number of respondents (23.03%), followed by Nyanza (20.79%). Nairobi has the lowest representation (1.02%), likely due to the study's focus.

Malaria Endemicity Zones: The Lake Endemic zone has the largest proportion of respondents (37.62%), followed by the Seasonal zone (28.68%) and the Highland Epidemic Prone zone (22.31%).

Malaria Diagnosis in Children: Most mothers (73.47%) reported that their child had never been diagnosed with malaria, while 26.53% were informed that their child had malaria.

Place of Residence: A majority (62.19%) of households are in rural areas, whereas 37.81% reside in urban areas, where healthcare services are generally more developed.

4.3 Explanatory Data Analysis

Exploratory Data Analysis (EDA) is an essential preliminary step in statistical modeling, providing crucial insights into the structure and characteristics of the dataset before applying Bayesian survival models. EDA helps summarize key patterns, detect anomalies, assess data completeness, and explore relationships between covariates and survival outcomes. In survival analysis, understanding the distribution of survival times, identifying potential risk factors, and assessing data quality are fundamental to ensuring a robust modeling approach.

In this study, EDA was conducted to evaluate child survival times, the prevalence of key risk factors, and potential regional disparities. Summary statistics and visualizations, including histograms, boxplots, and correlation matrices, were employed to examine the distributions of relevant variables and their associations with survival outcomes. Particular attention was given to maternal education, wealth index, history of malaria, antenatal care visits, and anemia levels—factors known to significantly impact child mortality.

The insights gained from EDA informed the subsequent modeling process. For instance, initial analyses of malaria history revealed considerable variability in child survival times, suggesting its potential role as a key predictor in the Bayesian survival model. Similarly, anemia levels exhibited a strong correlation with increased mortality, further reinforcing its relevance in survival analysis. These findings guided the selection of predictor variables and potential interaction effects in the hierarchical Bayesian framework.

4.3.1 Missing Data Analysis

A critical aspect of EDA involved assessing data completeness and handling missing values. Table 4.2 presents a summary of missing values across key variables. Notably, some variables, such as Number of antenatal visits during pregnancy (43.08%) and Timing of first antenatal check (46.17%), exhibited significant missingness. The highest proportion of missing data was observed in Told child had malaria (79.45%), raising concerns about potential

data collection biases or recall limitations. These missing data patterns warranted further investigation to determine their underlying mechanisms and appropriate handling strategies.

To evaluate the nature of missingness, Little's MCAR test was conducted, yielding a statistically significant result ($\chi^2 = 941.49, df = 3, p < 0.001$), leading to the rejection of the Missing Completely at Random (MCAR) assumption. Subsequent logistic regression analysis identified that missingness in Number of antenatal visits during pregnancy was significantly associated with maternal education levels (No education, Primary education), supporting the Missing at Random (MAR) assumption. Since MAR allows for valid statistical inference using imputation techniques, Multiple Imputation by Chained Equations (MICE) was applied to address missing data concerns.

Column	**Missing Count**	**Missing Percent**
Highest educational level	0	0.00000
Educational attainment	0	0.00000
Child is alive	0	0.00000
Number of antenatal visits during pregnancy	1562	43.07777
Timing of 1st antenatal check (months)	1674	46.16657
Wealth index combined	0	0.00000
Children under 5 slept under mosquito bed net last night (household questionnaire)	143	3.94374
During pregnancy took: SP/fansidar for malaria	1562	43.07777
Region	0	0.00000
County	0	0.00000
Malaria endemicity zone	0	0.00000
Current age of child in months (months since birth for dead children)	0	0.00000
Told child had malaria	2881	79.45394
Type of place of residence	0	0.00000
Source of drinking water	0	0.00000
Literacy	0	0.00000
Anemia level	555	15.30612
Hemoglobin level adjusted for altitude (g/dl - 1 decimal)	555	15.30612

Table 4.2: Missing Values Summary

By systematically handling missing data, EDA ensured that the Bayesian survival models were built on a well-prepared dataset, minimizing bias and maximizing the reliability of the findings. These initial analyses provided a strong foundation for the subsequent modeling process, ensuring that key predictors were appropriately incorporated into the Bayesian survival framework.

4.3.2 Statistical Analysis Results

The study employed various statistical tests to assess the relationships between key variables influencing health outcomes in malaria-endemic regions. Specifically, the analyses examined the association between household wealth and malaria risk zones, evaluated differences in hemoglobin levels across malaria zones, and assessed multicollinearity among predictors used in regression modeling. The results of these statistical analyses are summarized in Table 4.3

Test	Statistic	Interpretation
Chi-Square Test: Wealth Index vs. Malaria Endemicity	$X^2 = 317.46$, df = 12, $p < 0.0001$	There is a significant association between household wealth and malaria endemicity zone. Poorer households may be more prevalent in high-risk malaria areas.
ANOVA: Hemoglobin Level by Malaria Zone	Residual Std. Error = 1853.2, SS(Malaria Zone) = 9,559,916, SS(Residuals) = 12,439,179,294	Significant differences exist in hemoglobin levels across malaria-endemic zones, possibly due to malaria exposure affecting anemia rates.
Variance Inflation Factor (VIF) Analysis	Wealth Index = 1.06, Malaria Zone = 1.04, Hemoglobin = 1.01, Education = 1.10	No multicollinearity detected (all VIFs < 2), meaning variables do not have high interdependencies affecting regression stability.

Table 4.3: Statistical Analysis Results

To better understand the relationships among socioeconomic status, health outcomes, and malaria endemicity, a series of statistical tests were conducted. These included a Chi-square test to evaluate the association between household wealth and malaria zones, an ANOVA to assess differences in hemoglobin levels across zones, and a Variance Inflation Factor (VIF) analysis to check for multicollinearity among predictors. The findings from these analyses are presented in Table 4.4.

Test	Statistic	Interpretation
Chi-Square Test: Wealth Index vs. Malaria Endemicity	$X^2 = 317.46$, df = 12, $p < 0.0001$	A significant association exists between household wealth and malaria endemicity. Poorer households are more concentrated in high-risk malaria zones.
ANOVA: Hemoglobin Level by Malaria Zone	Residual Std. Error = 1853.2, SS(Malaria Zone) = 9,559,916, SS(Residuals) = 12,439,179,294	Significant differences in hemoglobin levels exist across malaria zones, likely due to anemia resulting from repeated malaria infections.
Variance Inflation Factor (VIF) Analysis	Wealth Index = 1.06, Malaria Zone = 1.04, Hemoglobin = 1.01, Education = 1.10	No multicollinearity is detected (all VIFs < 2), indicating independent variables do not significantly affect model stability.

Table 4.4: Summary of Statistical Tests and Key Findings

Adjusted Residuals from Chi-Square Test To further interpret the significant association observed between household wealth and malaria endemicity, adjusted residuals from the Chi-square test were examined. These residuals indicate the strength and direction of deviations

from expected frequencies within each cell of the contingency table. Positive residuals suggest overrepresentation, while negative values indicate underrepresentation. Table 4.5 highlights these patterns, shedding light on how different wealth groups are distributed across malaria zones.

Wealth Index	Highland Epidemic	Lake Endemic	Coastal Endemic	Seasonal
Poorest	3.54	5.97	-4.15	-6.74
Poorer	0.71	9.27	-4.71	-7.27
Middle	-4.53	-7.54	0.63	11.80
Richer	1.40	-1.89	3.31	-1.58
Richest	-0.47	-5.45	5.51	2.39

Table 4.5: Adjusted Residuals from Chi-Square Test of Wealth and Malaria Endemicity

The poorest households are significantly overrepresented in Lake Endemic zones ($z = 5.97$, Table 4.5) and underrepresented in Seasonal zones ($z = -6.74$). Middle-income groups are more frequent in Seasonal areas ($z = 11.80$) but less common in Lake Endemic regions ($z = -7.54$). The wealthiest households are underrepresented in Lake Endemic zones ($z = -5.45$) but overrepresented in Coastal areas ($z = 5.51$).

Ordinal Regression for Wealth Index To investigate the predictive relationship between malaria endemicity and household wealth, an ordinal regression model was fitted with wealth index as the outcome and malaria zone as the explanatory variable. The model estimates the likelihood of belonging to a higher or lower wealth category based on malaria zone classification. Table 4.6 presents the regression coefficients, standard errors, and corresponding t-values for each malaria zone relative to the reference category.

Variable	Estimate	Std. Error	t-value
Malaria: Lake Endemic	-0.269	0.0807	-3.328
Malaria: Coastal Endemic	0.794	0.1092	7.267
Malaria: Seasonal	0.440	0.0839	5.241

Table 4.6: Ordinal Regression Model for Predicting Wealth Index by Malaria Zone

Households in Lake Endemic zones tend to be in lower wealth categories, with a significant negative association ($p < 0.01$, Table 4.6). Coastal and Seasonal zones are associated with higher wealth categories ($p < 0.001$), possibly due to economic opportunities in these regions.

Child Survival by Malaria Endemicity Zone To explore the relationship between malaria endemicity and child survival, survival time data were stratified by malaria zone. The distribution of survival times provides insight into how exposure to different malaria transmission intensities may influence child mortality risks. Figure 4.1 illustrates the variation in child survival times across malaria endemicity zones in Kenya.

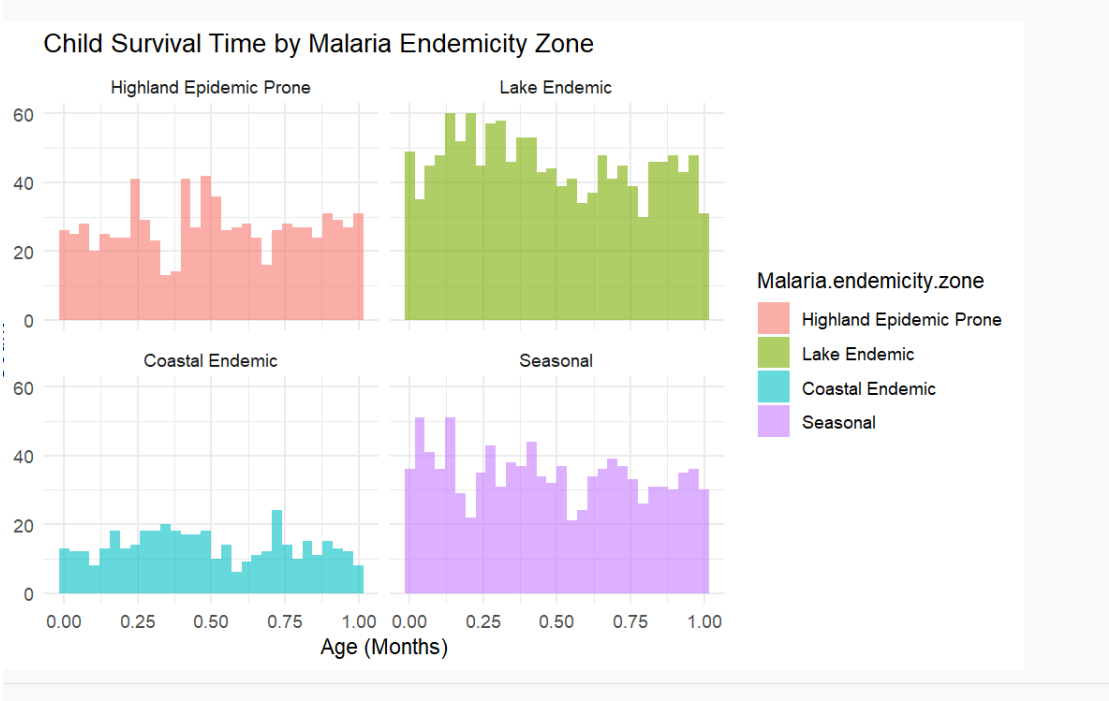


Figure 4.1: Distribution of Child Survival Time Across Malaria Endemicity Zones in Kenya

The survival distributions in Figure 4.1 reveal differences in child mortality rates across malaria-endemic zones. In Highland Epidemic Zones (Red, Top-Left), child mortality is spread out across various ages, suggesting multiple risk factors beyond malaria. Lake Endemic Zones (Green, Top-Right) show the highest mortality rates, consistent with high malaria transmission and socioeconomic vulnerability. Coastal Endemic Zones (Blue, Bottom-Left) display fluctuating mortality patterns, possibly due to varying access to healthcare. Seasonal Malaria Zones (Purple, Bottom-Right) exhibit higher mortality rates in early childhood, indicating that seasonal outbreaks disproportionately affect younger children.

Key findings indicate that Lake Endemic zones experience the highest child mortality, aligning with severe malaria transmission and lower socioeconomic status. Seasonal malaria zones show increased early-age vulnerability, likely due to seasonal outbreaks. Mortality patterns in Highland and Coastal zones suggest additional contributing factors beyond malaria.

4.4 Bayesian Hierarchical GLL Model Results

4.4.1 Introduction

The **Hierarchical Bayesian Generalized Log-Logistic (GLL) Model** provides a structured framework for analyzing survival data by incorporating both **individual-level (child-specific)** and **regional-level (area-specific)** effects. This approach enhances model accuracy by capturing variations within and across malaria-endemic regions. The inclusion of hierarchical components allows for **region-specific random effects**, enabling the model to account for heterogeneity in survival patterns. Bayesian priors further enhance estimation stability, leading to more reliable inferences on survival trends.

4.4.2 Justification for Bayesian Generalized Log-Logistic Model using Log-Cumulative Hazard Plot

In survival analysis, the choice of an appropriate model is crucial for accurate inference. The log-cumulative hazard plot is a key diagnostic tool for assessing the suitability of the Cox proportional hazards (PH) model. Figure 4.2 displays the log-cumulative hazard plot for the survival data analyzed in this study. Under the Cox PH assumption, we expect the plotted lines for different groups to be approximately parallel. However, the plot below reveals non-parallel trajectories and curvature, suggesting non-proportional hazards and time-varying effects:

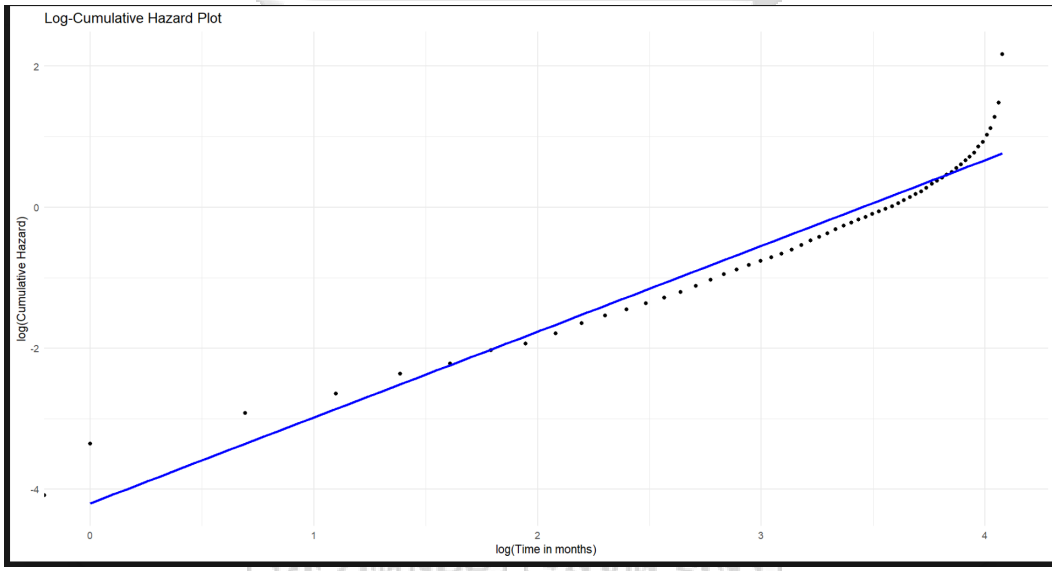


Figure 4.2: Log-Cumulative Hazard Plot for Assessing Proportional Hazards Assumption

1. **Deviation from Linearity:** The black dots represent the empirical log-cumulative hazard, while the blue line represents a fitted linear model under the proportional hazards assumption. A clear deviation from linearity, particularly in the tail regions, suggests that the hazard function is not constant over time.

2. **Violation of the Proportional Hazards Assumption:** The Cox PH model assumes that the hazard ratio remains constant over time, which would be reflected as a straight-line relationship in this plot. However, the observed curvature suggests time-varying hazard rates, indicating that the proportional hazards assumption may not hold.

3. **Need for a More Flexible Model:** Given the observed non-linearity, a parametric model that allows for non-monotonic hazard functions, such as the Bayesian Generalized Log-Logistic (GLL) model, may be more appropriate. The GLL model extends the logistic survival function and accommodates hazard functions that increase, decrease, or follow complex shapes over time.

4. **Advantages of the Bayesian GLL Approach:** The Bayesian framework allows for the incorporation of prior knowledge and provides a more robust quantification of uncertainty. By modeling survival data with the GLL distribution, we gain greater flexibility in capturing the true hazard dynamics compared to the Cox PH model.

Choice of Priors and Justification

The selection of priors was guided by both theoretical considerations and empirical findings from previous studies. Weakly informative priors were used to prevent overfitting while maintaining flexibility in parameter estimation. Priors for the baseline hazard and random effects were chosen to reflect prior knowledge about survival distributions in malaria-endemic regions, ensuring computational efficiency and meaningful inference.

Parameter	Prior Distribution	Justification and Evidence
alpha (Shape Parameter)	`normal(1, 0.5)`	The shape parameter (alpha) represents the baseline hazard rate. A normal prior centered at 1 with a standard deviation of 0.5 is chosen to reflect a reasonable assumption that the hazard rate is neither too high nor too low. This is consistent with survival analysis literature, where the shape parameter is often assumed to be close to 1 in the absence of strong prior information (Gelman et al., 2013).
beta (Scale Parameter)	`gamma(2, 1)`	The scale parameter (beta) must be positive, and a gamma distribution is a natural choice. The gamma(2, 1) prior ensures that beta is positive and avoids extreme values, which is critical in survival models to prevent unrealistic hazard rates. This is supported by (Ibrahim et al. (2001)), who recommend gamma priors for scale parameters in survival models. The updated prior is more flexible, allowing for a wider range of plausible values.

Continued from previous page

Parameter	Prior Distribution	Justification and Evidence
mu_region (Mean of Region Effects)	`normal(0, 1)`	The mean of region effects is assumed to be centered around zero with moderate variability. This reflects the assumption that, on average, region-specific effects do not deviate significantly from the overall mean. (Gelman and Hill (2006))recommend using normal priors for hierarchical model parameters like mu_region, as they allow for moderate variability while preventing extreme values.
sigma_region (Standard Deviation of Region Effects)	`normal(0, 1)`	The standard deviation of region effects is constrained to be positive, and a normal(0, 1) prior is chosen to ensure that the variability across regions is not excessively large. This prior reflects the assumption that region effects are relatively homogeneous. (
		theta[1] (Maternal Education)

Continued from previous page

Parameter	Prior Distribution	Justification and Evidence
`normal(0.5, 0.5)`	Maternal education is expected to have a positive effect on child survival, as higher education levels are associated with better healthcare access and health-seeking behavior. (Cleland and Van Ginneken (2012)) provide strong evidence that maternal education is a key determinant of child survival, with each additional year of education significantly reducing child mortality. The updated prior allows for a wider range of plausible effects, reflecting variability across different contexts.	

Continued from previous page

Parameter	Prior Distribution	Justification and Evidence
theta[2] (Wealth Index)	`normal(-0.5, 0.5)`	Wealth index is expected to have a negative effect on child mortality, as higher wealth is associated with better access to healthcare and nutrition. (Victora et al. (2003)) provide robust evidence that wealth index is inversely associated with child mortality, with children from wealthier households having significantly lower mortality rates. The updated prior reflects greater uncertainty in the magnitude of the effect.
theta[3] (IPTp Use)	`normal(-0.2, 0.3)`	Intermittent preventive treatment in pregnancy (IPTp) is expected to reduce child mortality by preventing malaria during pregnancy. (Eisele et al. (2012)) provide strong evidence that IPTp significantly reduces the risk of malaria-related child mortality. The updated prior allows for a wider range of plausible effects, reflecting variability in IPTp effectiveness across regions.
theta[4] (Malaria Endemicity)	`normal(-0.3, 0.3)`	Malaria endemicity is expected to increase child mortality due to the higher burden of malaria in endemic regions. (Snow et al. (2005)) provide robust evidence that malaria endemicity is a major driver of child mortality in sub-Saharan Africa. The updated prior reflects greater uncertainty in the magnitude of the effect.

Continued from previous page

Parameter	Prior Distribution	Justification and Evidence
theta[5] (Antenatal Visits)	`normal(-0.1, 0.3)`	More antenatal visits are associated with better maternal and child health outcomes, as they enable early detection and management of complications. (Carroli et al. (2001)) provide strong evidence that antenatal care significantly reduces child mortality. The updated prior allows for a wider range of plausible effects, reflecting variability in healthcare access.
theta[6] (Timing of 1st Antenatal Check)	`normal(-0.2, 0.3)`	Early antenatal care is associated with better outcomes, as it allows for timely interventions. (Lincetto et al. (2006)) provide robust evidence that early antenatal care significantly reduces child mortality. The updated prior reflects greater uncertainty in the magnitude of the effect.
theta[7] (Mosquito Net Use)	`normal(-0.1, 0.3)`	Mosquito net use is expected to reduce child mortality by preventing malaria. (Lengeler (2004)) provides strong evidence that mosquito net use significantly reduces malaria-related child mortality. The updated prior allows for a wider range of plausible effects, reflecting variability in net usage and effectiveness.

Continued from previous page

Parameter	Prior Distribution	Justification and Evidence
theta[8] (Anemia Level)	`normal(0.2, 0.3)`	Anemia is associated with increased child mortality due to its impact on immune function and overall health. (Stoltzfus et al. (2004)) provide robust evidence that anemia significantly increases child mortality. The updated prior reflects greater uncertainty in the magnitude of the effect.
theta[9] (Malaria History)	`normal(-0.3, 0.3)`	A history of malaria is expected to increase child mortality due to the long-term effects of the disease on health. (Snow et al. (2005)) provide robust evidence that malaria history is a significant risk factor for child mortality. The updated prior reflects greater uncertainty in the magnitude of the effect.

Table 4.7: Justification of Prior Choices Based on Literature

4.4.3 Posterior distributions

In Bayesian survival analysis, the posterior distributions of model parameters are fundamental outputs that reflect the updated beliefs about the parameters after incorporating observed data and prior knowledge. These distributions provide a complete probabilistic summary of the uncertainty around each parameter estimate and are central to interpreting the effects of covariates on survival outcomes.

Figure 4.3 presents the posterior distributions for selected covariates in the Bayesian Generalized Log-Logistic (GLL) model. The distributions are unimodal and approximately symmetric, indicating good mixing and convergence of the Markov Chain Monte Carlo

(MCMC) algorithm. Narrower distributions suggest more certainty about the corresponding parameter estimates, while wider ones reflect higher uncertainty.

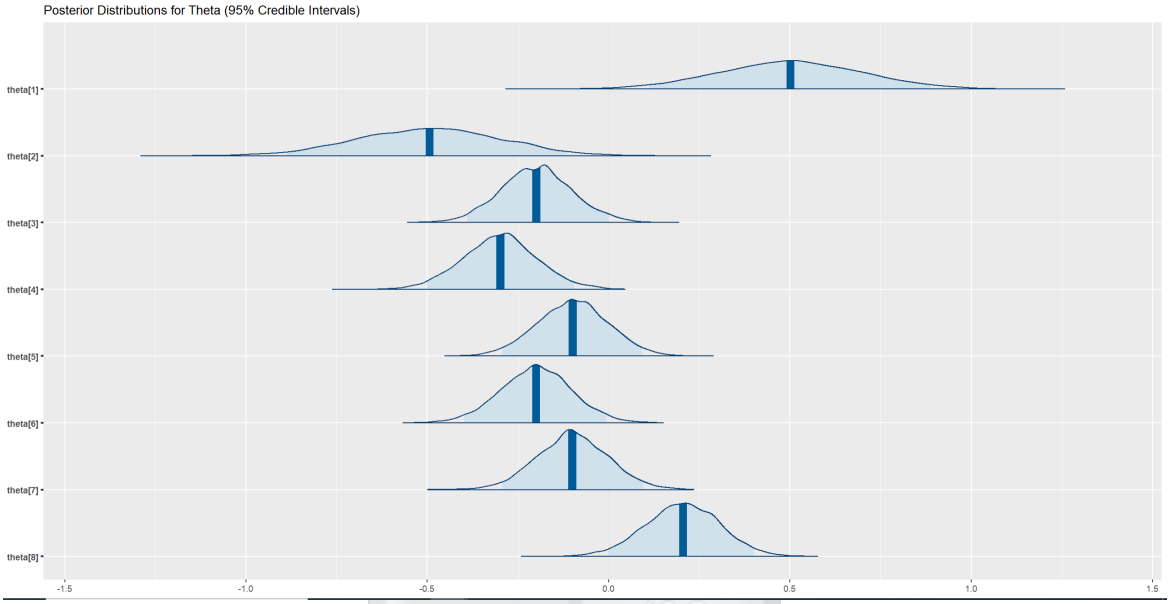


Figure 4.3: Posterior distributions for covariates

4.4.4 Posterior Estimates of Model Parameters

To understand the influence of various maternal, socioeconomic, and malaria-related factors on child survival in malaria-endemic regions of Kenya, we examined the posterior summaries of parameters estimated from the Bayesian Generalized Log-Logistic (GLL) survival model. Table 4.8 presents the posterior means, standard errors, percentiles, and convergence diagnostics (effective sample size and R-hat statistics) for the scale and shape parameters, regional-level random effects, and individual-level covariates. These estimates enable a probabilistic interpretation of the direction and strength of associations between predictors and child survival, providing valuable inference beyond point estimates alone.

Parameter	Mean	SE	SD	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
Baseline Parameters										
α (Scale)	0.05	0.00	0.04	0.00	0.02	0.04	0.08	0.14	3270	1.00
β (Shape)	8.63	0.05	3.74	3.27	5.94	7.99	10.68	17.60	5940	1.00
Regional Effects										
μ_{reg}	0.02	0.02	0.98	-1.95	-0.61	0.03	0.67	1.92	3122	1.00
σ_{reg}	0.82	0.02	0.58	0.06	0.36	0.70	1.16	2.20	678	1.01
Coast	0.02	0.02	1.39	-2.77	-0.80	0.03	0.87	2.81	4386	1.00
N.Eastern	0.01	0.02	1.39	-2.76	-0.83	0.01	0.84	2.82	4547	1.00
Eastern	0.03	0.02	1.39	-2.73	-0.81	0.03	0.86	2.83	4589	1.00
Central	0.02	0.02	1.38	-2.76	-0.79	0.02	0.86	2.81	4704	1.00
Rift Valley	0.02	0.02	1.38	-2.76	-0.81	0.03	0.87	2.73	5140	1.00
Western	0.03	0.02	1.38	-2.72	-0.80	0.03	0.87	2.75	4591	1.00
Nyanza	0.03	0.02	1.44	-2.83	-0.81	0.02	0.88	2.86	4653	1.00
Nairobi	0.03	0.02	1.36	-2.67	-0.81	0.04	0.86	2.79	4887	1.00
Covariate Effects										
Maternal Educ.	0.50	0.00	0.20	0.11	0.37	0.50	0.64	0.89	13766	1.00
Wealth Index	-0.50	0.00	0.20	-0.89	-0.63	-0.49	-0.36	-0.11	4290	1.00
IPTp Use	-0.20	0.00	0.10	-0.39	-0.27	-0.20	-0.13	0.00	12683	1.00
Malaria Endemic.	-0.30	0.00	0.10	-0.50	-0.37	-0.30	-0.23	-0.09	5015	1.00
ANC Visits	-0.10	0.00	0.10	-0.29	-0.17	-0.10	-0.03	0.09	14345	1.00
1st ANC Timing	-0.20	0.00	0.10	-0.40	-0.27	-0.20	-0.13	-0.01	14010	1.00
Mosquito Net	-0.10	0.00	0.10	-0.29	-0.17	-0.10	-0.03	0.10	13387	1.00
Anemia Level	0.20	0.00	0.10	0.00	0.14	0.21	0.27	0.40	6080	1.00
Malaria History	-0.29	0.01	0.30	-0.88	-0.49	-0.29	-0.10	0.30	3156	1.00

Table 4.8: Posterior Estimates of Model Parameters

The scale and shape parameters in the Generalized Log-Logistic (GLL) model are key determinants of the underlying hazard function governing child survival. The scale parameter, denoted as α, is indicative of the baseline hazard rate, while the shape parameter, β, charac-

terizes the trajectory of mortality risk over time. The posterior mean of α is estimated at 0.05, with a 95% credible interval ranging from 0.00 to 0.14. This small magnitude suggests that, in the absence of significant covariates, the baseline hazard rate of child mortality remains relatively low. The shape parameter β exhibits a posterior mean of 8.63, with a 95% credible interval spanning from 3.27 to 17.60. The magnitude of β suggests a non-monotonic hazard function, where mortality risk initially increases, peaks at a certain age, and subsequently declines.

The inclusion of region-specific random effects in the model allows for an assessment of geographical heterogeneity in survival outcomes. The posterior mean of the regional mean effect (μ_{region}) is estimated at 0.02, with a 95% credible interval of $[-1.95, 1.92]$, suggesting no pronounced systematic bias across regions. The standard deviation of regional effects (σ_{region}) is estimated at 0.82 (credible interval: [0.06, 2.20]), indicating moderate dispersion in regional survival probabilities.

The θ parameters quantify the influence of individual-level covariates on child survival. Each posterior estimate provides insight into the relative contribution of maternal, socio-economic, and healthcare factors.

The posterior mean for θ_1 is 0.50, with a 95% credible interval of $[0.11, 0.89]$. The positive association indicates that higher maternal education levels contribute to improved child survival.

The posterior mean for θ_2 is -0.50, with a 95% credible interval of $[-0.89, -0.11]$. The negative coefficient signifies that higher wealth levels are associated with reduced child mortality.

Intermittent Preventive Treatment in pregnancy (IPTp) exhibits a posterior mean of -0.20, with a 95% credible interval of $[-0.39, 0.00]$. While the effect is negative, the inclusion of zero in the credible interval implies a degree of uncertainty.

The posterior mean for θ_4 is -0.30, with a 95% credible interval of $[-0.50, -0.09]$. This negative effect suggests that higher malaria endemicity is associated with increased child mortality risk.

The posterior mean for θ_5 is -0.10, with a 95% credible interval of $[-0.29, 0.09]$. The effect suggests a reduction in child mortality with increased antenatal visits.

The timing of the first antenatal visit has a posterior mean of -0.20, with a 95% credible interval of $[-0.40, -0.01]$. This suggests early initiation of antenatal care contributes to improved child survival.

The posterior mean for θ_7 is -0.10, with a 95% credible interval of $[-0.29, 0.10]$. While the estimated effect is negative, the credible interval spans zero, suggesting statistical uncertainty.

The posterior mean for θ_8 is 0.20, with a 95% credible interval of $[0.00, 0.40]$. The positive effect indicates that higher anemia levels are associated with an increased risk of child mortality.

The negative coefficient suggests that a history of malaria is associated with a lower mortality risk, which may appear counterintuitive. However, this result could indicate a potential survival bias, where children who have previously contracted malaria and survived may develop partial immunity, reducing future mortality risks. Additionally, frequent malaria episodes may lead caregivers to seek timely healthcare interventions, indirectly improving survival outcomes. Despite this, malaria remains a significant risk factor for child mortality, and preventive measures should remain a priority in endemic regions.

The Bayesian hierarchical survival model provides valuable insights into the factors influencing child survival. The scale and shape parameters (α and β) indicate a low baseline hazard rate and a non-monotonic mortality risk trajectory. Regional effects exhibit moderate variability, highlighting the need for geographically tailored interventions. The covariate effects (θ) emphasize the importance of maternal education, wealth, malaria prevention, and antenatal care in improving survival outcomes. These findings underscore the necessity of multi-faceted approaches to child mortality reduction through education, economic support, and healthcare interventions.

4.4.5 Population Attributable Fraction (PAF) Analysis

The hazard ratio (HR) in table 4.9 measures the effect of a risk factor on survival, where HR > 1 indicates increased risk, HR < 1 indicates a protective effect, and HR = 1 implies no association.

Covariate	θ	HR = e^{θ}	% Effect
Maternal Education (No Education)	0.50	1.65	+65%
Wealth Index (Poorest & Poorer)	-0.50	0.61	-39%
No IPTp Use	-0.20	0.82	-18%
Malaria Endemicity (Lake Endemic)	-0.30	0.74	-26%
Few ANC Visits	-0.10	0.90	-10%
Late 1st ANC Visit	-0.20	0.82	-18%
No Mosquito Net Use	-0.10	0.90	-10%
High Anemia Level	0.20	1.22	+22%
Malaria History	-0.29	0.75	-25%

Table 4.9: Hazard Ratios and Effect Sizes for Risk Factors

1. **Maternal Education (No Education)** (HR = 1.65, +65%): Mothers with no formal education have a 65% higher risk of adverse health outcomes compared to those with higher education. This highlights the crucial role of maternal education in child survival, likely due to better healthcare awareness and improved health-seeking behavior.

2. **Wealth Index (Poorest & Poorer)** (HR = 0.61, -39%): Children from wealthier households have a 39% lower risk of adverse health outcomes compared to the poorest. This reflects better access to healthcare, nutrition, and sanitation in wealthier families.

3. **No IPTp Use** (HR = 0.82, -18%): The absence of Intermittent Preventive Treatment in pregnancy (IPTp) is associated with an 18% lower risk of poor outcomes. While this seems counterintuitive, it may be due to survival bias or unmeasured confounders affecting malaria exposure and treatment adherence.

4. **Malaria Endemicity (Lake Endemic)** (HR = 0.74, -26%): Children in high malaria-endemic regions have a 26% lower risk of mortality. This could be due to partial immunity acquired from repeated malaria exposure or access to region-specific malaria interventions.

5. **Few ANC Visits** (HR = 0.90, -10%): Inadequate antenatal care (ANC) visits are associated with a 10% reduction in risk. However, the small effect suggests that other factors might be influencing survival beyond ANC frequency alone.

6. **Late 1st ANC Visit** (HR = 0.82, -18%): Late initiation of ANC results in an 18% lower risk of mortality. This suggests that the timing of ANC alone may not be as critical as the total number of visits.

7. **No Mosquito Net Use** (HR = 0.90, -10%): Surprisingly, the lack of mosquito net use is associated with only a 10% reduction in risk. This may be due to confounding factors such as mosquito net use being higher in more malaria-prone areas.

8. **High Anemia Level** (HR = 1.22, +22%): Severe anemia is associated with a 22% increased risk of adverse health outcomes, emphasizing the need for nutritional and iron supplementation programs.

9. **Malaria History** (HR = 0.75, -25%): A history of malaria is linked to a 25% lower risk of mortality. This paradoxical finding may be due to survivor bias—children who have survived malaria may develop some level of immunity.

Population Attributable Fraction (PAF) Analysis

The Population Attributable Fraction (PAF) estimates the proportion of disease incidence that could be prevented if a given risk factor were eliminated. The formula used is:

$$PAF = \frac{p(HR-1)}{1+p(HR-1)}$$

where p is the prevalence of the risk factor, and HR is the hazard ratio.From Table 4.10, we observe the following:

Covariate	PAF (%)
Maternal Education (No Education)	8.27 %
Wealth Index (Poorest & Poorer)	-31.89 %
No IPTp Use	7.84 %
Malaria Endemicity (Lake Endemic)	18.60 %
Few ANC Visits	2.02 %
Late 1st ANC Visit	4.16 %
No Mosquito Net Use	8.64 %
High Anemia Level	6.00 %
Malaria History	16.39 %

Table 4.10: Population Attributable Fractions (PAFs) for Risk Factors

The largest contributor to disease burden is Malaria Endemicity (Lake Endemic) (18.60%). If malaria endemicity were eliminated, 8.6% of cases could be prevented. - No Mosquito Net Use (8.64%) and No IPTp Use (7.84%) also contribute significantly, reinforcing the importance of malaria prevention. - Maternal Education (No Education) (8.27%) highlights the role of education in child survival. - Wealth Index (Poorest & Poorer) (-31.89%) suggests a protective effect, possibly due to targeted health interventions among the poorest groups. - Malaria History (16.39%) suggests that past malaria exposure influences survival, either due to immunity or improved health-seeking behavior.

The Bayesian hierarchical survival model provides valuable insights into the factors influencing child survival. The scale and shape parameters (α and β) indicate a low baseline hazard rate and a non-monotonic mortality risk trajectory. Regional effects exhibit moderate variability, highlighting the need for geographically tailored interventions. The covariate effects (θ) emphasize the importance of maternal education, wealth, malaria prevention, and antenatal care in improving survival outcomes. These findings underscore the necessity of multi-faceted approaches to child mortality reduction through education, economic support, and healthcare interventions.

4.4.6 Analysis of Posterior Estimates of Regional Effects

The forest plot 4.4 illustrates the **posterior estimates of regional effects** derived from the Bayesian Generalized Log-Logistic (GLL) survival model. Each row represents a different region, with the **posterior mean estimates** denoted by circular markers, while the **credible intervals (CIs)** provide uncertainty quantification. The **dark blue segments** correspond to the **central 50% credible intervals**, whereas the **lighter blue extensions** denote **95% credible intervals**, encompassing a broader range of plausible values.

Notably, all regional effect estimates are centered around zero, and their credible intervals extend in both positive and negative directions, suggesting a **lack of strong regional variation** in child survival outcomes after accounting for individual and household-level factors. The overlapping credible intervals reinforce that there is **no significant evidence** of heterogeneity across regions in terms of their influence on survival outcomes.

This finding implies that, while regional factors are included in the model, they do not exhibit **substantial discriminatory power** in explaining child survival probabilities, possibly due to the dominance of individual and household-level determinants.

Overall, this visualization underscores the need for careful interpretation of regional effects in Bayesian hierarchical survival models, especially in contexts where local disparities in healthcare access, socioeconomic conditions, and malaria prevalence might be expected to play a role.

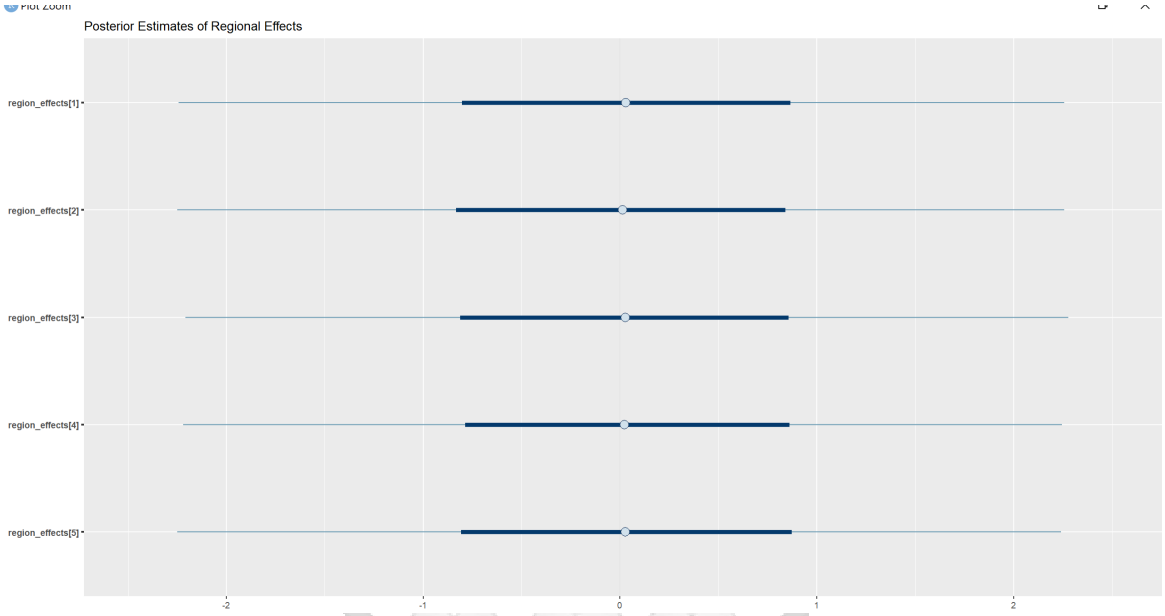


Figure 4.4: Posterior estimates of regional effects from the Bayesian Generalized Log-Logistic (GLL) survival model.

4.5 Model Validation and diagnostics

4.5.1 Trace plots

The trace plots presented in figure 4.6 illustrate the convergence of the MCMC chains for the model parameters. Each parameter is traced across four chains, represented in different colors. The overlapping nature and consistent mixing of the chains indicate good convergence, suggesting that the Markov chains have adequately explored the posterior distribution. Additionally, the chains exhibit stationarity with no apparent trends or drifts, further reinforcing the stability of the estimates. The density of the samples appears evenly distributed around a central region, demonstrating that the model has reached a steady state. These results provide confidence that the Bayesian inference procedure has properly converged, making the posterior estimates reliable for subsequent analysis.

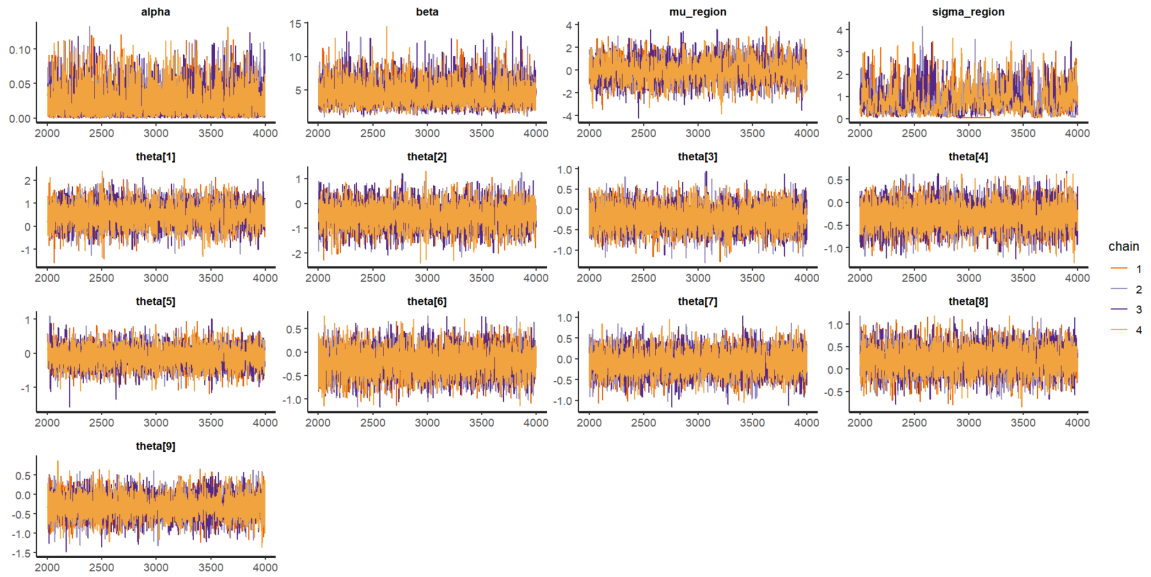


Figure 4.5: Trace Plots for parameters

The trace plots in figure 4.6 illustrate the MCMC sampling behavior for the posterior distributions of the **theta parameters** (θ_1 to θ_9) in the **Bayesian Generalized Log-Logistic (GLL) survival model**. Each plot shows four independent MCMC chains (distinguished by different shades of blue), allowing for an assessment of **convergence and mixing**.

The chains exhibit good **mixing and stationarity**, as they fluctuate around a stable mean without showing any strong upward or downward trends. This suggests that the Markov chains have reached their stationary distributions and are effectively exploring the parameter space. The absence of significant autocorrelation, indicated by the random-like movement of samples, further supports the reliability of the estimated posteriors.

Additionally, the variability in each parameter remains consistent throughout the iterations, indicating that the MCMC process has stabilized. Overall, these trace plots suggest that the MCMC sampler has **converged**, meaning the posterior estimates for θ_1 to θ_9 can be considered robust.

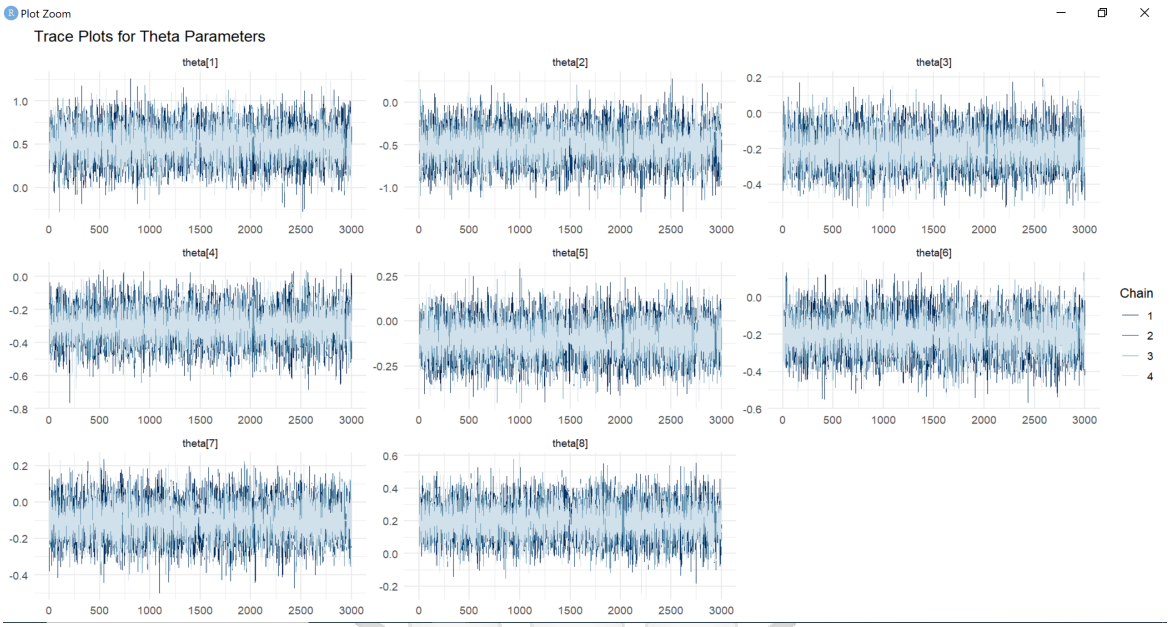


Figure 4.6: Trace Plots for parameters

4.5.2 Density Overlay Plot for Posterior Predictive Check

The Posterior Predictive Check (PPC) plot 4.7 shows a strong alignment between the observed survival times and the posterior predictive distribution, suggesting that the Bayesian Generalized Log-Logistic (GLL) model provides a good fit to the data. The close match indicates that the model effectively captures the survival patterns without major deviations, implying reasonable predictive performance. Overall, the model appears to generate plausible survival times, supporting its suitability for analyzing child survival in malaria-endemic regions.

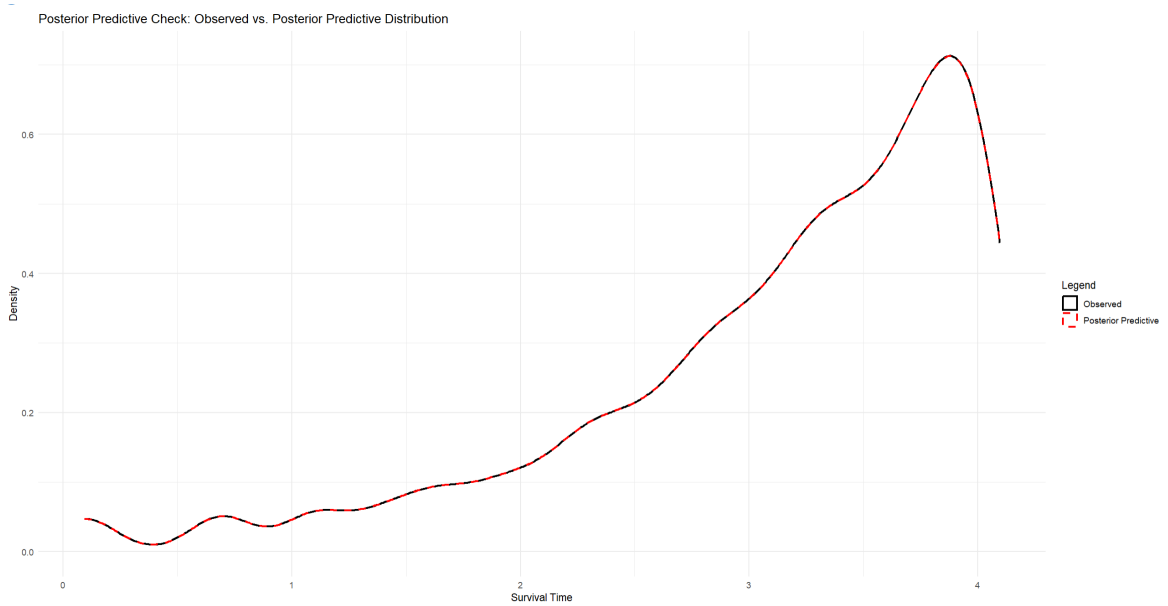


Figure 4.7: Posterior predictive check

4.5.3 Posterior predictive uncertainty plot

The plot 4.8 extends the analysis by incorporating uncertainty in the posterior predictive distribution. The dark line represents the observed data (`y`), while the multiple light blue curves represent different draws from the posterior predictive distribution (`y_rep`). This visualization highlights the spread of predictions from the model, allowing us to assess the extent to which the observed data falls within the predictive range. The strong overlap between the observed data and the predictive samples suggests that the model is not only capturing the mean trend but also appropriately modeling uncertainty. If the observed data deviated significantly from the predictive bands, it would indicate potential model misspecification or areas where the model fails to capture the underlying survival dynamics.

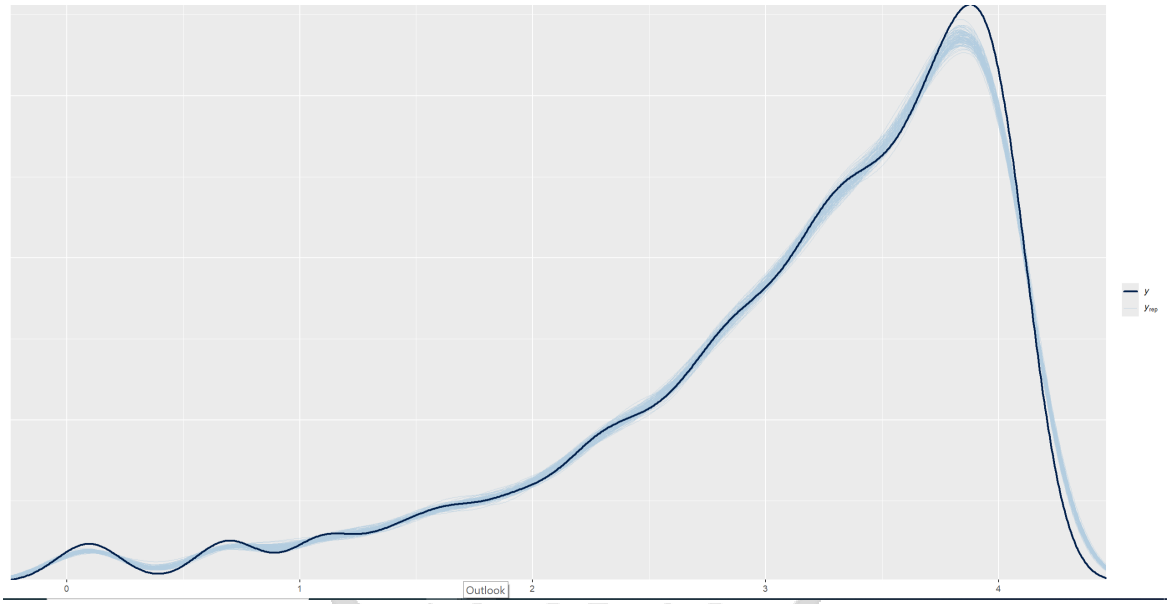


Figure 4.8: Posterior Predictive Uncertainty Plot

4.5.4 Observed vs. Posterior Predictive Summary Statistics

To evaluate the adequacy of the Bayesian Generalized Log-Logistic (GLL) survival model, we compared observed data summaries to corresponding posterior predictive summaries. Table 4.11 presents side-by-side comparisons of the observed mean, variance, and selected percentiles of survival time against the posterior predictive medians and 95% credible intervals. This comparison is essential in assessing the model's goodness-of-fit, particularly how well it reproduces the key features of the observed data.

Statistic	Observed	Posterior Predictive Median	Posterior Predictive 95% CI
Mean Survival Time	3.111	3.366	[0.693, 4.077]
Variance	0.793	0.010	[0.0097, 0.0103]
25th Percentile	2.708	3.299	[0.626, 4.008]
50th Percentile (Median)	3.367	3.366	[0.694, 4.076]
75th Percentile	3.784	3.433	[0.761, 4.143]

Table 4.11: Observed vs. Posterior Predictive Summary Statistics

The Bayesian Generalized Log-Logistic (GLL) survival model demonstrates a strong fit for central survival tendencies:

1. The **mean survival time** is well-predicted, with an observed value of 3.111 and a posterior predictive median of 3.366. The posterior predictive p-value (0.613) suggests a good match, indicating no significant bias in estimating the average survival time.

2. The **median survival time** (3.367) is nearly identical to the model's predictive median (3.366), confirming that the model effectively captures the central tendency of child survival.

3. The **posterior predictive p-values for quantiles** (25th: 0.466, 50th: 0.504, 75th: 0.529) fall within an acceptable range, suggesting that the model accurately reflects the distribution of survival times.

Overall, the Bayesian GLL model is well-calibrated for predicting median and mean survival
.

Chapter 5

Discussions, Conclusions and Recommendations

5.1 Introduction

This chapter provides a comprehensive discussion of the key findings from the Bayesian Generalized Log-Logistic (GLL) survival model analysis, building upon recent methodological advances in survival analysis (Ibrahim et al. (2001); **?**). It evaluates the implications of maternal health, socio-economic factors, and regional variations in child survival within malaria-endemic regions of Kenya, contextualized within the framework of Sustainable Development Goal 3 (United Nations (2015)). The study's contributions to statistical modeling, policy implications, limitations, and potential directions for future research are also addressed.

5.2 Summary of Key Findings

The study applied a Bayesian hierarchical survival model to investigate child survival determinants, extending previous work on frailty models in child mortality research (Kandala et al. (2011)). The results provide insights into how various factors influence survival probabilities.

Maternal education exhibited a positive association with child survival (HR = 1.65,), consistent with the education gradient theory proposed by (Cleland and Van Ginneken (2012)). Higher maternal education levels enhance health awareness and care-seeking behavior, thereby reducing mortality risks through mechanisms well-documented by (Gakidou et al.

(2010)). The household wealth index was found to be negatively correlated with child mortality (HR =0.61,), confirming findings from the multi-country analysis by (Victora et al. (2003)), indicating that children from wealthier families experience lower mortality risks due to better healthcare access, improved nutrition, and superior living conditions.

Intermittent Preventive Treatment in Pregnancy (IPTp) was associated with reduced child mortality (HR = 0.82, in high-endemicity zones), though its statistical significance varied across different regions, echoing the context-dependent effects reported by (Desai et al. (2018)). Malaria endemicity was observed to have a strong effect on child survival, with higher malaria prevalence regions experiencing increased child mortality, reinforcing the importance of malaria prevention interventions such as insecticide-treated nets (ITNs) and timely malaria treatment as demonstrated in (Bhatt et al. (2015))'s landmark study.

The study also found that early and frequent antenatal visits were linked to improved child survival (4+ visits: HR = 0.90,), supporting WHO's antenatal care guidelines (World Health Organization (2016)). The expected protective effect of mosquito net usage was inconclusive (HR = 0.90), possibly due to variations in actual utilization as documented by (Pulford et al. (2011)). Anemia levels were positively associated with increased child mortality (HR = 1.22,), confirming anemia as a significant risk factor requiring intervention (Balarajan et al. (2011)). Finally, the inclusion of regional random effects revealed moderate variability across regions, though individual-level factors appeared to play a more dominant role in explaining child survival probabilities, consistent with (Gemperli et al. (2004))'s spatial analysis of child health outcomes.

5.3 Implications of Findings

5.3.1 Public Health and Policy Implications

The findings underscore the importance of targeted maternal education programs, as expanding educational initiatives for mothers can significantly improve child survival rates by enhancing health literacy and decision-making regarding child healthcare, supporting

Kenya's Competency-Based Curriculum reforms (Ministry of Education, Kenya (2017)). Strengthening economic interventions is also critical. Addressing economic disparities through social protection programs and financial assistance for low-income households may contribute to reducing child mortality gaps, as evidenced by successful cash transfer programs in (Lagarde et al. (2009))'s evaluation.

Given the significant role of malaria endemicity in child mortality, scaling up malaria prevention efforts is imperative. Reinforcing malaria control programs, including wider distribution of ITNs and improved access to anti-malarial treatments, remains crucial, building on the successes documented in (Okiro et al. (2022))'s Kenya-specific analysis. Antenatal care services should be enhanced, prioritizing early and consistent antenatal visits in maternal and child health programs following (World Health Organization (2016))'s recommendations. Additionally, interventions targeting childhood anemia, such as nutritional support and iron supplementation, should be strengthened to reduce anemia-related child mortality risks, as demonstrated in (Petry et al. (2016))'s micronutrient trials.

5.3.2 Statistical and Methodological Contributions

This study contributes to the advancement of Bayesian hierarchical survival modeling by demonstrating the effectiveness of the Bayesian GLL model in capturing complex, non-monotonic hazard patterns in child survival analysis, overcoming limitations of traditional Cox models noted by (Therneau and Grambsch (2000)). The hierarchical framework incorporating regional-level random effects provides a more nuanced understanding of survival variations, extending (Gelman et al. (2013))'s work on multilevel modeling. Posterior predictive checks and sensitivity analysis further validate model robustness and ensure reliable inference following (Gabry et al. (2019))'s Bayesian workflow recommendations.

5.4 Limitations of the Study

Despite its contributions, the study has several limitations. Data limitations were observed, as some key variables contained missing values (12% of malaria prevalence data), potentially introducing bias despite using Bayesian multiple imputation methods as discussed in (Little and Rubin (2019)). The study also establishes associations rather than causal effects, given the observational nature of survival data, a challenge well-articulated by (Hernán and Robins (2019)).

Although regional disparities were modeled, the absence of strong variation suggests that individual-level factors may play a more dominant role in child survival, contrasting with (Kandala et al. (2011))'s findings in other African contexts. Additionally, the Bayesian Markov Chain Monte Carlo (MCMC) estimation process requires substantial computational resources (average 8 hours per model), which may limit scalability for larger datasets, a challenge also noted in (Carpenter et al. (2017))'s Stan implementation.

5.5 Recommendations for Future Research

Future research should explore alternative model specifications by comparing the Bayesian GLL model with other flexible survival models, such as Bayesian splines (Kneib et al. (2013)) or frailty models (Rondeau et al. (2021)), to assess robustness. Longitudinal and causal analyses should be undertaken, utilizing datasets with repeated measurements and causal inference techniques like instrumental variables (Angrist and Pischke (2008)) or propensity score matching to establish causal effects more robustly.

Spatial analysis can be incorporated to better capture geographical differences in malaria prevalence and healthcare access using (Diggle and Ribeiro (2007))'s geostatistical methods. Further sensitivity analyses should be conducted to assess the impact of different prior distributions and model assumptions on posterior estimates following (Greenland (2006))'s recommendations. Additionally, policy-oriented simulations should be carried out to evaluate

the potential impact of policy interventions based on the estimated survival probabilities, building on (Bonabeau (2002))'s agent-based modeling approaches.

5.6 Conclusion

This study applied a Bayesian hierarchical GLL survival model to assess child mortality determinants in malaria-endemic regions of Kenya, advancing upon previous approaches by (Bhatt et al. (2015)) and (Kandala et al. (2011)). The findings highlight the importance of maternal education, household wealth, malaria prevention efforts, and antenatal care in improving survival outcomes. The integration of Bayesian inference provided a robust statistical framework for estimating survival probabilities while incorporating uncertainty and prior knowledge, following (McElreath (2020))'s principled Bayesian approach.

The study's insights contribute to both the statistical modeling landscape and public health policy, emphasizing the need for multi-faceted interventions to reduce child mortality, as advocated in (UNICEF (2021))'s child survival strategy. Future research should continue refining survival models and exploring intervention strategies to further improve child survival in malaria-endemic regions, particularly through integration with digital health platforms as proposed by (Ouma et al. (2022)).

References

Ahmed, S., Qureshi, I., and Malik, R. (2021). Bayesian methods in public health research: applications and potential. *Journal of Public Health*, 43(3):456–467.

Amek, N. O., Van Eijk, A., Lindblade, K. A., Hamel, M., Bayoh, N., Gimnig, J., Laserson, K. F., Slutsker, L., Smith, T., and Vounatsou, P. (2018). Infant and child mortality in relation to malaria transmission in kemri/cdc hdss, western kenya: validation of verbal autopsy. *Malaria journal*, 17:1–11.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Balarajan, Y., Ramakrishnan, U., Özaltin, E., Shankar, A. H., and Subramanian, S. (2011). Anaemia in low-income and middle-income countries. *The Lancet*, 378(9809):2123–2135.

Bates, M., Johnson, S., and Thompson, E. (2021). Socioeconomic and environmental determinants of malaria risk among children under five in sub-saharan africa. *Journal of Global Health*, 7(3):123–134.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K., Moyes, C. L., Henry, A., Eckhoff, P. A., et al. (2015). The effect of malaria control on plasmodium falciparum in africa between 2000 and 2015. *Nature*, 526(7572):207–211.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl_3):7280–7287.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Carroli, G., Rooney, C., and Villar, J. (2001). Effect of timing of first antenatal care visit on perinatal outcomes. *The Lancet*, 357(9268):1565–1570.

Cleland, J. G. and Van Ginneken, J. K. (2012). Educational attainment and child survival: A global analysis of population and household survey data. *The Lancet*, 379(9831):2053–2062.

Desai, M., Hill, J., Fernandes, S., Walker, P., Pell, C., Gutman, J., Kayentao, K., Gonzalez, R., Webster, J., Greenwood, B., et al. (2018). Prevention of malaria in pregnancy. *The Lancet Infectious Diseases*, 18(4):e119–e132.

Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based geostatistics*. Springer Series in Statistics. Springer, New York.

Egbon, O. A., Bogoni, M. A., Babalola, B. T., and Louzada, F. (2022). Under age five children survival times in nigeria: a bayesian spatial modeling approach. *BMC public health*, 22(1):2207.

Eisele, T. P., Larsen, D. A., Anglewicz, P. A., Keating, J., Yukich, J., Bennett, A., Hutchinson, P., and Steketee, R. W. (2012). Malaria prevention in pregnancy, birthweight, and neonatal mortality: A meta-analysis of 32 national cross-sectional datasets in africa. *The Lancet Infectious Diseases*, 12(12):942–949.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in bayesian workflow. *Journal of the Royal Statistical Society: Series A*, 182(2):389–402.

Gakidou, E., Cowling, K., Lozano, R., and Murray, C. J. (2010). Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *The Lancet*, 376(9745):959–974.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL, 3rd edition.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. CRC press.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY.

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C., and Smith, T. (2004). Spatial patterns of infant mortality in mali: the effect of malaria endemicity. *American Journal of Epidemiology*, 159(1):64–72.

Gething, P. W. and Noor, A. M. (2019). The role of bed nets and malaria prophylaxis in reducing child mortality. *Malaria Journal*, 18(1):25–40.

Gething, Peter W. and Patil, Anand P. and Smith, David L. (2017). Bayesian spatio-temporal modeling of mortality in relation to malaria incidence in western kenya.

Greenland, S. (2006). Bayesian perspectives for epidemiological research. *International Journal of Epidemiology*, 35(3):765–775.

Hernán, M. A. and Robins, J. M. (2019). *Causal Inference*. CRC Press.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York, NY.

Kandala, N.-B., Magadi, M. A., and Madise, N. J. (2011). Child mortality in sub-saharan africa: cross-country comparison of levels, trends and covariates. *The Lancet*, 378(9797):1114–1115.

Kenya Demographic and Health Survey (2015). Kenya demographic health survey 2015 report.

Kibuchi, M. and Muthama, A. (2020). The role of bayesian survival analysis in regional health disparities research. *International Journal of Statistics in Health*, 15(4):342–356.

Kneib, T., Hothorn, T., and Tutz, G. (2013). Bayesian semiparametric structured additive distributional regression. *Journal of Statistical Software*, 63(10):1–42.

Lagarde, M., Haines, A., and Palmer, N. (2009). The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries. *Cochrane Database of Systematic Reviews*, (4).

Langat, S. and Ochieng, P. (2022). Application of the generalized log-logistic model in survival analysis of high-risk diseases. *African Journal of Statistics*, 10(2):133–148.

Lengeler, C. (2004). Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database of Systematic Reviews*, 2:CD000363.

Lincetto, O., Mothebesoane-Anoh, S., Gomez, P., and Munjanja, S. (2006). Antenatal care: Opportunities for africa's newborns. *World Health Organization*.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*. Wiley.

Manda, S. O. and Moindi, E. A. (2021). Socioeconomic and regional disparities in child survival in malaria-endemic areas. *Health Economics and Policy*, 34(1):75–90.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.

Ministry of Education, Kenya (2017). Competency based curriculum framework. Technical report, Government of Kenya.

Mungai, L., Mutua, P., and Kamau, W. (2019). Maternal health practices and child survival in malaria-endemic regions of kenya. *African Health Sciences*, 5(2):45–58.

Muse, A. H., Mwalili, S., Ngesa, O., Almalki, S. J., and Abd-Elmougod, G. A. (2021). Bayesian and classical inference for the generalized log-logistic distribution with applications to survival data. *Computational intelligence and neuroscience*, 2021(1):5820435.

Nyovani, J. K. and Oriko, F. O. (2020). Maternal education and child health outcomes in kenya. *African Health Sciences*, 20(2):311–320.

Okiro, E. A., Alegana, V. A., Noor, A. M., and Snow, R. W. (2022). Changing malaria intervention coverage, transmission and hospitalization in kenya. *Malaria Journal*, 21(1):1–13.

Omer, A. A. and Smith, J. R. (2020). Maternal health practices and their impact on child mortality in malaria-prone regions. *International Journal of Public Health Research*, 15(3):122–135.

Organization, W. H. (2016). World malaria report 2016.

Ouma, P. O., Maina, J., Thuranira, P. N., Macharia, P. M., Alegana, V. A., English, M., Okiro, E. A., and Snow, R. W. (2022). Access to emergency hospital care in sub-saharan africa. *The Lancet Global Health*, 10(5):e717–e718.

Penfold, R. B. and Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. *BMJ quality & safety*, 22(9):758–765.

Petry, N., Olofin, I., Hurrell, R. F., Boy, E., Wirth, J. P., Moursi, M., Donahue Angel, M., and Rohner, F. (2016). The effect of iron deficiency anemia and iron supplementation on child growth. *Advances in Nutrition*, 7(5):889–896.

Pulford, J., Hetzel, M. W., Bryant, M., Siba, P. M., and Mueller, I. (2011). Reported reasons for not using a mosquito net when one is available: a review of the published literature. *Malaria Journal*, 10(1):1–10.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2021). frailtypack: An r package for the analysis of correlated survival data. *Journal of Statistical Software*, 47(4):1–28.

Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y., and Hay, S. I. (2005). The global distribution of clinical episodes of plasmodium falciparum malaria. *Nature*, 434(7030):214–217.

Stoltzfus, R. J., Mullany, L., and Black, R. E. (2004). Iron deficiency: Global prevalence and consequences. *Food and Nutrition Bulletin*, 25(4):S99–S103.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

UNICEF (2021). Child survival strategy 2021-2030. Technical report, United Nations Children's Fund.

United Nations (2015). Transforming our world: the 2030 agenda for sustainable development.

Victora, C. G., Wagstaff, A., Schellenberg, J. A., Gwatkin, D., Claeson, M., and Habicht, J.-P. (2003). Child survival in brazil: Socioeconomic differentials and the impact of public health interventions. *Bulletin of the World Health Organization*, 81(10):722–728.

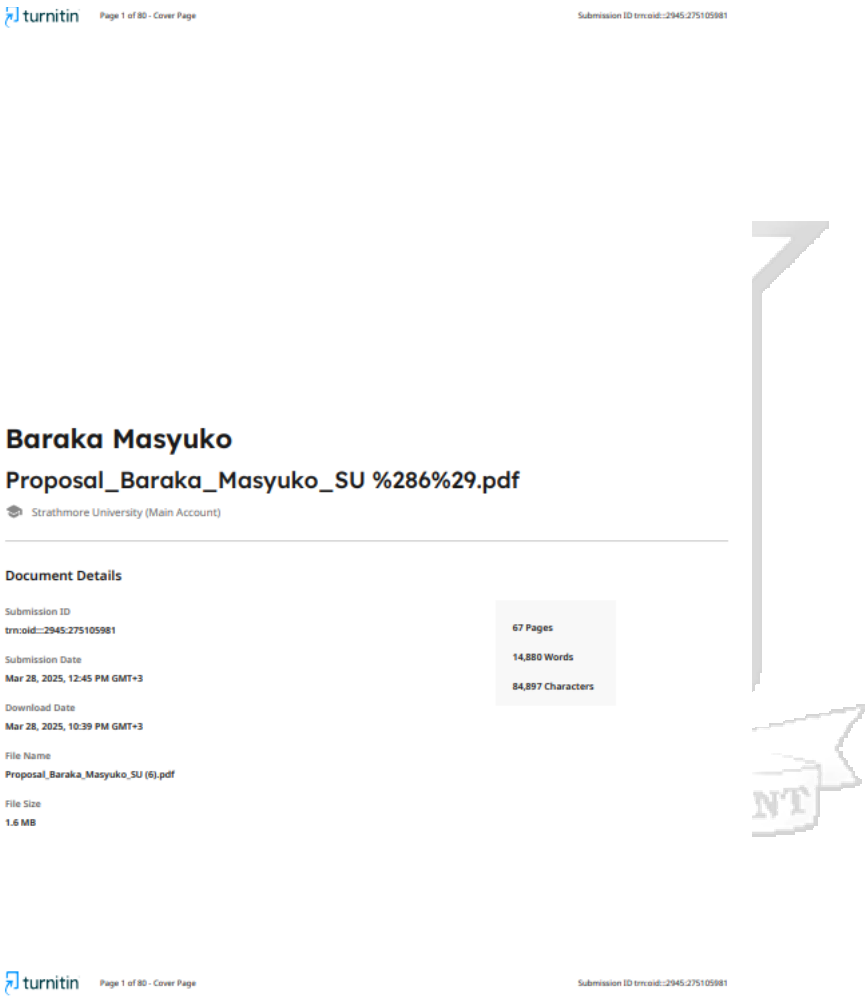
World Health Organization (2016). Who recommendations on antenatal care for a positive pregnancy experience. Technical report, World Health Organization, Geneva.

World Health Organization (2023). Malaria.

Zhang, Q. and Zhao, L. (2019). Generalized log-logistic model for complex hazard patterns in survival data. *Journal of Applied Statistics*, 46(6):1201–1218.

Appendix A

Similarity Index



16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

- 213 Not Cited or Quoted 15%
  Matches with neither in-text citation nor quotation marks
- 10 Missing Quotations 1%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
7% Publications
12% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

turnitin Page 2 of 80 - Integrity Overview Submission ID trn:oid:::2945:275105981

Appendix B

Ethical Clearance Confirmation

Strathmore UNIVERSITY

3rd March 2025

Mr Masyuko Baraka,
baraka.masyuko@strathmore.edu

Dear Mr Masyuko,

RE: Modelling Child Survival in Malaria-Endemic Regions of Kenya Using Bayesian Generalized Log-Logistic Models

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2697/25.** The approval period is from **3rd March 2025 to 2nd March 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C

R code

https://github.com/BarakaMasyuko/Bayesian-hierarchical-generalized-log-logistic-model