Assessing Latent Factors In Malaria Health Behaviors And Access To Preventive Care Using Frailty Models

Mwangi, Henry Njenga

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

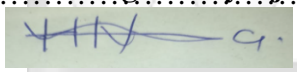
Nairobi, Kenya

June 2025

Declaration

I declare that this work has not been previously submitted and approved for 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 dissertation itself.

Name: **Mwangi Henry Njenga**

Signature: 

Date: May 27, 2025

Approval

The dissertation of Mwangi Henry Njenga was reviewed and approved by the following:

Dr. Evans Otieno Omondi

Supervisor,

Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu

Dean,

Institute of Mathematical Sciences, Strathmore University.

Prof. Bernard Shibwabo

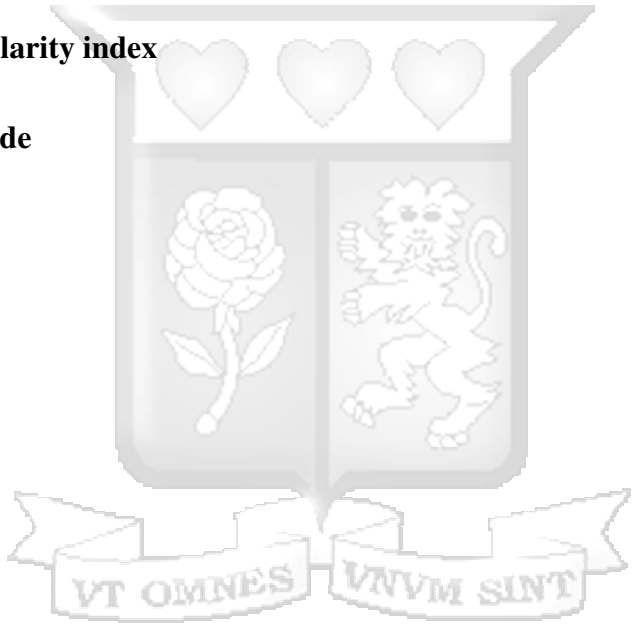
Director,

Office of Graduate Studies, Strathmore University.

Abstract

Malaria remains a major public health challenge in sub-Saharan Africa, particularly among children under five. Despite progress in prevention, disparities in adoption persist due to socioeconomic, geographic, and individual factors. This study examines the influence of unobserved household and cluster-level factors on malaria prevention behaviors and access to care in Kenya using inverse Gaussian frailty models. Secondary data from the 2020 Kenya Demographic and Health Survey (KDHS) is analyzed to assess the impact of malaria-endemic zones, digital connectivity, and child-specific factors on preventive care adoption. A semi-parametric inverse Gaussian frailty model quantifies latent variability at household and cluster levels, providing deeper insights into malaria prevention and treatment behaviors. Results show that frailty models significantly outperform the Cox model (ΔAIC > 18,000), with gamma frailty performing best in high-variability settings ($\theta > 0.2$, ΔAIC = 7.68). Mobile phone access strongly predicts ITN adoption (HR ≈ 1.21, $p < 0.001$), while internet use and social media have no effect ($p > 0.70$). Children in endemic areas adopt ITNs earlier (Estimate = 0.233, $p < 0.001$), with moderate frailty variance ($\theta = 0.006$). Younger children exhibit 3.4× higher preventive care uptake ($\beta = -3.368$, $p < 0.001$), with high frailty variance ($\theta = 0.5$), indicating substantial unobserved heterogeneity. Digital connectivity shows minimal cluster effects ($\theta = 0.005$), suggesting uniform access. These findings highlight the need for targeted mobile health strategies, age-specific prevention programs, and frailty-adjusted methodologies to enhance malaria control interventions.

Table of contents



List of figures

List of tables

List of abbreviations

AIC	Akaike Information Criterion	AFT	Accelerated Failure Time
ACT	Artemisinin-Based Combination Therapy	ANC	Antenatal Care
ART	Antiretroviral Therapy	BIC	Bayesian Information Criterion
CAR frailty	Conditional Autoregressive Frailty	CD4	Cluster of Differentiation 4 (a type of spatial frailty model)
DHS	Demographic and Health Survey	IST	Intermittent Screening and Treatment
ITN	Insecticide-Treated Net	KDHS	Kenya Demographic and Health Survey
KNBS	Kenya National Bureau of Statistics	LRT	Likelihood Ratio Test
MICE	Multiple Imputation by Chained Equations	PMM	Predictive Mean Matching
SMS	Short Message Service	TB	Tuberculosis
WHO	World Health Organization		

Acknowledgement

First and above all, I thank the Almighty God for the knowledge, strength and good health during the study period.

I am grateful to my parents, brothers and sisters for their immeasurable support throughout the study period.

I am indebted to my supervisors, Dr. Evans Otieno Omondi for his availability, kind guidance, and support without which this proposal would not have been a success.

I am grateful to my classmates for their support and encouragement.

Dedication

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

Chapter 1

Introduction

1.1 Background to the study

Malaria remains a significant worldwide public health challenge , most particularly in sub-Saharan Africa, where endemic zones experience high morbidity and mortality rates among vulnerable populations in continuity, especially children under the age of five (World Health Organization, 2023). Despite there being advances in preventive measures such as use of insecticide-treated nets (ITNs) and antimalarial drugs, there is still substantial variability that persists in the adoption and adherence to these interventions across different regions, households, and individuals. This variability is often attributed to a complex interplay of socioeconomic, geographic, and individual factors that influence health behaviors and access to preventive care . Having a good Understanding of the latent, unobserved factors that contribute to these disparities is very crucial for designing targeted interventions that can effectively reduce malaria transmission and improve health outcomes (Vanderslott, 2019).

1.1.1 Geographic factors and malaria endemicity

Endemic zones are pronounced to having Geographic variation in malaria prevalence and behavior patterns,in shaping malaria-related health behaviors and access to preventive care factors such as regional climate, ecological conditions, and healthcare infrastructure play a central role (Hennessee, 2022). Previous studies have indicated that individuals residing in endemic areas are particularly often more aware of malaria prevention methods but face difficulties in accessing resources (Alam et al., 2016). These studies have often overlooked on latent factors which are unmeasured variables that represent underlying social or envi-

ronmental influences that could further explain regional disparities in malaria prevention behaviors. Using advanced statistical approaches, such as frailty models, which provides an opportunity to account for unobserved heterogeneity at the household or cluster level, capturing the hidden variability across malaria-endemic and non-endemic zones (Camilleri et al., 2017).

1.1.2 Digital connectivity and health behaviors

Rapid growth of digital technology,access to mobile phones ,internet access and use of social media has brought significant change in health communication and engagement taking care of the low-resource settings (Fitzpatrick, 2023). The use of Mobile technology and social media have shown promise and indication of increasing awareness and facilitation of access to health information, which has shown to positively impact health behaviors (Free et al., 2013). Looking at digital connectivity at a malaria perspective , it has shown that it helps in enhancing preventive behaviors by ensuring households have access to information on prevention of malaria and possible treatment options which are timely (Smith, 2023). However, there is limited research on how access to these digital resources directly influences household malaria prevention behaviors, such as the use of ITNs or adherence to antimalarial treatment protocols. Examining the role of mobile technology and social media exposure in influencing health behavior adoption through inverse Gaussian frailty models can help identify underlying digital barriers or facilitators to malaria prevention.

1.1.3 Child-specific factors in malaria prevention and treatment

Under five year old children are at the highest risk of malaria infection and complications which makes them a critical target population for targeted interventions (Organization, 2022). Several and different Child-specific factors, such as age, malaria symptoms, and access to artemisinin-based combination therapy (ACT), highly influence preventive care measures and treatment adherence (Banek et al., 2014). Younger children, are expected to be more reliant on caregiver knowledge and access to healthcare, while older children may have completely

different behavioral and treatment needs. Understanding how these child-specific factors contribute to disparities in malaria prevention and treatment uptake requires approaches that can account for the unobserved heterogeneity among individuals and clusters. Incorporating frailty models, particularly those with an inverse Gaussian distribution, allows researchers to quantify the influence of these latent factors on child-specific outcomes, potentially uncovering differential treatment adherence or preventive care adoption patterns based on unobserved household or cluster-level characteristics (Duchateau and Janssen, 2008).

1.1.4 Frailty models in public health research

In public health research frailty models have emerged as a crucial and powerful tool. They have enabled researchers to be able to account for unobserved heterogeneity in health behavioral and survival outcomes. Frailty models help in introducing a latent variable also well known as the frailty which helps represent the influence of unobserved risk factors that are shared within clusters which include households or regions (Therneau et al., 2003; Wienke, 2003). Inverse Gaussian frailty models are well suited for analysis that has significant unobserved heterogeneity which in turn allows for flexibility in modelling complex dependencies in clustered data (Gutierrez, 2002). Application of the Inverse Gaussian frailty models aims to capture hidden variability in health behaviors that are malaria related across different geographic zones ,assess the digital connectivity impact on adoption of health behaviors and exploring how child specific factors have an impact on preventive care and adherence to treatment.

1.2 Statement of the problem

Malaria remains as the leading cause of illness and death in sub-Saharan Africa, disproportionately affecting children under five and pregnant women in malaria-endemic regions. While preventive interventions such as insecticide-treated nets (ITNs), antimalarial drugs, and indoor residual spraying have proven effective, their impact is hindered by disparities in

adoption and adherence. These disparities, evident across geographic, socioeconomic, and demographic dimensions, suggest the presence of unmeasured, latent factors that influence malaria-related health behaviors and access to care. Despite significant research on malaria prevention, most studies have primarily focused on measurable determinants, such as income, education, and geographic location, with limited attention to unobserved heterogeneity at the household and cluster levels. These latent factors representing shared social, environmental, or structural influences have remained inadequately explored, most particularly in endemic versus non-endemic zones. In order to address this gap, it requires thorough use of advanced statistical models that are capable of quantifying and accounting for this unobserved and hidden variability.

In addition, the rapid expansion and developmental growth in digital connectivity, including mobile phone access, internet use, and social media exposure, brings forth an opportunity to enhance health behaviors through improved access to malaria prevention information and services. However, little is known about how digital connectivity influences household decisions regarding malaria prevention. Understanding this relationship is critical to leveraging technology for public health interventions.

Children under five, a particularly vulnerable group, face unique challenges in accessing malaria prevention and treatment. Factors such as age, the presence of malaria symptoms, and access to artemisinin-based combination therapy (ACT) play critical roles in determining their health outcomes. However, unmeasured household or cluster-level influences may further affect these outcomes, necessitating methods that can address both measured and unmeasured variability. Given these challenges, there is a need for robust modeling approaches to investigate how unobserved factors affect malaria-related health behaviors, particularly in endemic zones. The use of inverse Gaussian frailty models offers an effective means to capture and quantify this hidden variability, hence it enables a deeper understanding of dynamics on malaria prevention dynamics.

This study addresses these gaps by applying semi-parametric inverse Gaussian frailty models to examine the influence of endemicity, digital connectivity, and child-specific factors on malaria prevention behaviors and access to care in Kenya. By integrating these approaches,

the research aims to inform targeted, data driven interventions that reduce disparities and improve malaria prevention outcomes.

1.3 Research objectives

1.3.1 General objective

The main objective of this study is to assess the influence of unobserved household and cluster-level factors on malaria prevention behaviors and access to care in malaria-endemic regions, using semi-parametric inverse Gaussian frailty models in Kenya.

1.3.2 Specific Objectives

1. To develop and apply inverse Gaussian frailty models to quantify the impact of malaria-endemic zones on malaria-related health behaviors and access to preventive care at household and cluster level.

2. To assess the impact of digital connectivity on the adoption of malaria prevention measures by Investigating the influence of mobile phone access , internet use , and social media exposure on household health behaviors.

3. To apply inverse Gaussian frailty models to quantitatively analyze the influence of child age, malaria symptoms, and ACT treatment on the likelihood of receiving malaria preventive care.

1.4 Justification of the study

Due to the persistent challenges and disparities in malaria prevention and treatment uptake in sub-Saharan Africa, most particularly in malaria endemic regions in Kenya. Despite having progress in control of malaria through interventions like ITNs (Insecticide treated

nets) and ACT (Artemisinin-Based Combination Therapy.), there has been variability in adoption and adherence remains due to socioeconomic, geographic, and individual factors (World Health Organization, 2023). The 2020 KDHS data provides a comprehensive well defined dataset to help in exploring these disparities and identifying latent, unobserved factors that are affecting health behaviors at the household and cluster levels. Employing semi-parametric inverse Gaussian frailty models will help in enabling a deeper, clear and thorough understanding of unobserved heterogeneity in behaviors across endemic and non-endemic zones. Additionally, it will help in assessing the contribution of digital connectivity (mobile phones, internet, and social media) in promotion of malaria prevention measures that addresses research of public health in digital age which is a growing area. This study's insights will inform targeted, context-specific interventions to improve access and adherence to malaria prevention resources, especially for vulnerable populations like children under five, who remain disproportionately affected.

1.5 Significance of the study

This research addresses crucial gaps in understanding the unobserved and hidden household- and cluster-level factors influencing malaria prevention behaviors and access to care in Kenya, a region with substantial disparities in malaria burden. Utilizing the 2020 KDHS data and applying advanced inverse Gaussian frailty models, the research aims to uncover latent variability across endemic and non-endemic zones, thereby providing insights into the hidden drivers of health behaviors. In addition, it evaluates the transformative potential of digital connectivity in regards to mobile phone access, internet use, and social media exposure in improving the adoption of malaria preventive measures. Furthermore, by integrating child-specific variables such as age, symptoms, and ACT treatment adherence, this study seeks to inform targeted interventions for vulnerable populations, particularly children under five. The findings will contribute to evidence-based policies aimed at reducing malaria-related morbidity and mortality in Kenya.

Chapter 2

Literature Review

2.1 Introduction

The literature review gives a comprehensive and thorough dissection and exploration of key areas that are important to malaria prevention and control in Sub-Saharan Africa, focusing on Kenya. It starts by contextualizing on the persistent malaria burden, most particularly among the vulnerable populations such as children under five and pregnant women, reflecting on the disease's impact socioeconomically. The Geographic factors are examined, emphasizing their important role in healthcare disparities and disease transmission. Behavioral and Socioeconomic determinants are thoroughly explored, hence addressing challenges and inhibitors to preventive measures like antimalarial adherence and ITN use. The transformative potential and ability of digital connectivity in enhancing health behaviors is well discussed alongside child-specific vulnerabilities in malaria prevention. The review also introduces frailty models as advanced statistical tools to address unobserved heterogeneity, underscoring their relevance to the study. Finally, it identifies critical research gaps, particularly in understanding latent geographic and socioeconomic factors, setting the stage for the current investigation into equitable and effective malaria control strategies.

2.2 The burden of malaria in sub-saharan africa

Malaria continues to exert a significant burden on Sub-Saharan Africa, attributing to high mortality and morbidity rates, particularly among children who are under five and pregnant women. The World Health Organization (World Health Organization, 2023) reports that Sub-Saharan Africa accounts for over 95 percent of global malaria cases and deaths, with

children under five comprising approximately 80 percent of fatalities. This vulnerability is largely due to their immature immune systems and limited access to effective preventive measures and treatments. The malaria economic toll in the region is substantial. Sarma et al. (2019) estimate that endemic countries collectively lose billions annually in direct healthcare costs, decreased productivity, and impaired economic growth. Households that are affected by malaria will often have catastrophic health expenditures, that further exacerbates the poverty cycles in resource-limited settings.

In Kenya, malaria remains a great public health challenge despite the crucial and significant investments in control measures. High-burden areas, that include western Kenya and the coastal regions, experience consistently high and elevated prevalence rates due to suitable climatic conditions for mosquito breeding (Kenya National Bureau of Statistics , KNBS). Efforts in combating the disease, including the distribution of insecticide-treated nets (ITNs) and artemisinin-based combination therapies (ACTs), have achieved varying levels of success. For instance, while ITN coverage has expanded, consistent use remains a challenge due to factors such as cultural misconceptions and lack of access (Organization, 2022). Furthermore, resistance to antimalarial drugs and insecticides poses a growing threat to progress in these regions.

Beyond health systems, the malaria persistence is influenced by broader socioeconomic and environmental factors. For instance, the limited access to quality healthcare in the rural areas, in conjunction with inadequate infrastructure, significantly affects early diagnosis and treatment (Okumu et al., 2022). Moreover, climatic variability, such as increased rainfall, often leads to spikes in transmission, overwhelming local health facilities.

Despite these many challenges, there has been regional initiatives such as the Zero Malaria Starts with Me campaign and international support from the Global Fund that have catalyzed the progress in malaria control. Achieving sustained reductions in malaria prevalence needs addressing systemic challenges and barriers, that include improving healthcare access, reducing economic inequities, and integrating innovative solutions like digital health tools to enhance prevention and treatment adherence (Okumu et al., 2022).

2.3 Geographic factors and malaria-endemic zones

Geographic factors play an important role in determining the malaria distribution and intensity in transmission, as shown by studies focusing on environmental conditions, altitudes, and regional malaria trends. Malaria transmission is highly sensitive to environmental factors which include humidity, rainfall, and temperatures. Areas with higher temperatures, particularly those between 20°C and 30°C, are considered an ideal ground for breeding more specifically for the Anopheles mosquitoes, which is considered the primary vector for malaria. In addition, areas that receive rainfall consistently and water sources, such as wetlands or irrigation systems, further accelerate the spread of malaria since they provide breeding sites for these mosquitoes (Leal Filho et al., 2023; McMahon et al., 2021) .

Topography also plays a large role in influencing the incidence of malaria. Looking at a case where highland areas which are considered to be cooler will experience lower malaria prevalence, but this can change with variation in climate and also fluctuation in seasons. Taking and considering Ethiopia, it has been seen that outbreaks of malaria are very common in areas considered to be lowlands, but highland regions with variable altitudes can experience sporadic malaria transmission due to seasonal factors and local environmental conditions (McMahon et al., 2021). Similarly, regions with significant changes in elevation , like the Choke Mountains in Ethiopia, present varying malaria patterns, where transmission of malaria can happen in lower altitudes but not in higher elevations (McMahon et al., 2021). In addition, location geographically determines the temporal patterns of malaria. In many sub-Saharan African countries, such as Kenya and Ethiopia, the onset of malaria cases is often associated to particular seasons, that are highly impacted by both local geography and climatic conditions. These patterns reflect the importance of geography in structuring the risk and burden of malaria (Leal Filho et al., 2023).

Geographic factors, that including climate, topography, and the availability of water, are crucial to understanding malaria endemicity. These factors have an association with regional health systems and malaria control efforts to shape the intensity and distribution of malaria

cases across different areas. Effective malaria control requires tailored strategies that account for the geographic and climatic characteristics of each region.

2.4 Socioeconomic and behavioral determinants of malaria prevention

Socioeconomic and behavioral factors have a crucial role in shaping behaviors in malaria prevention and access to interventions. Economic disparities are considered to be a major determinant, with households considered to be lower-income they are mostly unable to afford insecticide-treated nets (ITNs) or timely access to malaria treatments. Subsidized or free distribution of ITNs has been shown to significantly increase use and ownership, particularly among vulnerable populations. However, consistent and persistent usage remains a hinderance due to the many misconceptions and knowledge gaps, such as beliefs about the safety of ITNs during pregnancy or mistrust of antimalarial drugs (Hill et al., 2013).

Behavioral determinants, that include knowledge, attitudes, and perceptions about severity of malaria and prevention efficacy, have a strong influence on preventive behaviors. Looking at one of the Studies in Nigeria, the study highlights that perceived response efficacy and self-efficacy were important key factors driving the uptake of malaria prevention behaviors, including the consistent and persistent ITN usage. This was observed particularly in rural and urban Nigerian communities, where awareness and confidence in the effectiveness of ITNs were linked to increased usage rates. Such psychosocial factors are critical in designing effective malaria prevention campaigns (Duodu et al., 2022). Change of Social behavior campaigns leveraging the ideation model have addressed these challenges by enhancing and ensuring there is awareness and motivation for behaviors in prevention (Hill et al., 2013). Public engagement and education that are tailored to local contexts are important to overcome challenges which include the fear of side effects and cultural misconceptions. Several Strategies that include integration of malaria education into antenatal care (ANC) visits have proven to be very important in the increase of ITN usage and adherence to

preventive drug regimens like intermittent preventive treatment in pregnancy (Hill et al., 2013; Monroe et al., 2021).

2.5 The role of digital connectivity in health behaviors

The rapid advancement and growth of digital technology which includes mobile phones, access to internet, and social media has profoundly helped in reshaping communication in health, particularly in settings which are constrained in terms of resources. Digital platforms have become powerful tools for promoting health behaviors, including those targeting prevention in malaria.

In the study (Fitzpatrick, 2023) highlights the effectiveness of mobile technologies in helping to improve access to health information, demonstrating how SMS reminders and mobile apps have increased the use of insecticide-treated nets (ITNs) and adherence to antimalarial treatments. Such interventions have significantly enhanced preventive behaviors in malaria-endemic regions. For example, initiatives discussed by NetHope have successfully utilized mobile platforms to disseminate educational content and enable real-time surveillance of malaria hotspots (NetHope, 2024; World Health Organization (WHO), 2024). Similarly, (Free et al., 2013) emphasized on the potential of digital health tools in fostering change in behavior. The findings suggested that interventions leveraging mobile and internet platforms effectively raise awareness and influence behaviors, especially when tailored to the needs of underserved populations. The World Health Organization (WHO) also supports the integration of digital solutions into malaria surveillance programs, further demonstrating their role in improving health outcomes (MIT Solve, 2024; NetHope, 2024). In another study, (Smith, 2023) delved into the specific influence of digital strategies on malaria prevention, noting that access to timely, accurate information via mobile technology empowers households to adopt preventive measures and seek appropriate treatments. However, the study also underscored the need for further research to quantify the direct influence of digital connectivity on behaviors like ITN usage and treatment adherence. The use of statistical models, such as inverse Gaussian

frailty models, has been proposed to explore unobserved factors affecting health behaviors, providing a deeper understanding of digital barriers and facilitators (NetHope, 2024).

Despite these advancements, there are still challenges that are still persistent. Digital divides, particularly in rural and low-income regions, limit the reach and effectiveness of these interventions. The WHO and NetHope stress the importance of addressing disparities in mobile phone ownership and internet access to ensure equitable benefits from digital health tools (MIT Solve, 2024; NetHope, 2024) .

Digital connectivity offers significant promise in transforming health behaviors and supporting malaria prevention efforts. However, overcoming access and literacy barriers is essential to fully harness its potential.

2.6 Child-specific determinants in malaria prevention

The under five-year-old children are highly predisposed and susceptible to malaria, this makes them a primary focus for strategies on treatment and also on prevention. There are several factors that are child specific and have an influence on the outcomes, which include age, symptom recognition, and regimen treatment adherence such as artemisinin-based combination therapies (ACTs). Banek et al. (2014) puts more emphasis that children who are of lower age are heavily dependent on caregivers in the symptoms recognition and ACT administration. This can lead to delayed or inconsistent treatment. Similarly, challenges in dosing have been highlighted in studies, which have revealed that under-dosing in children, even under controlled settings, compromises efficacy in treatment and leads to an increase in drug resistance risk (Group, 2013). There have been Efforts to improve on prevention of malaria for children which puts more focus on health interventions in the community, this helps promote early detection and consistent treatment. The WHO shows the role of integrated community-based programs in improving access and adherence to ACTs, particularly in rural and underserved areas (Organization, 2022). Additionally, advanced modeling techniques like inverse Gaussian frailty models are being employed to explore

hidden household-level factors influencing treatment outcomes, providing deeper insights into disparities in malaria care.

2.7 Frailty models

Frailty models are a group of statistical tools which are used to cater for heterogeneity that is unobserved in data, to be very specific during analysis of survival times and occurrences. Considering these models are very impactful in analysis of survival data, these models help in modelling the dependence between observations in clusters which include families, hospitals and also other groupings. Frailty models are mostly often used in several fields, that include sociology, economics and healthcare.The most important feature of the frailty model is including a latent variable. This is known as frailty which represents factors that are unobserved which influence event of interest. This helps to account for random effects or individual heterogeneity may not be captured by observable covariates.

There are several approaches which are used in the use of frailty models, which includes shared frailty models, the frailty term is assumed to be common across members of a group or cluster. This is mostly useful when analyzing data coming from studies considered to be multi-centered, or when subjects who are in the same group are likely to have unobserved characteristics that are similar. An important development in the field has been the introduction of various frailty distributions, such as the gamma, inverse Gaussian, and lognormal distributions, which have different assumptions about the underlying dependence structure of the data (Gorfine and Zucker, 2023; Khalil and Gobbens, 2023).

2.7.1 Frailty models in public health research

Frailty models have now become very necessary tools in research conducted in public health they are most particular helpful in handling heterogeneity that is unobserved in survival data and clustered data structures. These models help in accounting for the availability of latent factors that impact outcomes in health, they mostly help in providing results that are reliable

and more accurate which are better than models used traditionally, particularly when dealing with dependent observations. The frailty term inclusion, helps in representing variables that are unmeasured variables which influence outcomes of health or survival, this helps in allowing for modeling of disparities in health and the effects of interventions across groups. In public health research, frailty models have been widely used to examine various issues in health, which include diseases considered infectious and chronic diseases, as well as in behaviors in health. A study conducted in South West Ethiopia has utilized shared frailty models to help in exploring the time to death and the factors associated with tuberculosis (TB) patients. The study helped in demonstrating that frailty models are very effective in looking for heterogeneity that is unobserved, such as socio-economic and environmental factors, that may impact outcomes in treatment. This approach revealed that some patients were more susceptible to failure treatment due to these factors that are unmeasured, this highlights on the need of latent variables being considered in outcomes of health analysis (Jabir et al., 2022).

Frailty models have had a wide application in HIV/AIDS research to address the unobserved heterogeneity and clustering effects, most specifically in survival analysis. A study that was conducted at Debre Tabor Referral Hospital, Ethiopia, used the Accelerated Failure Time (AFT) models in conjunction with gamma and inverse Gaussian frailty distributions to help in analyzing the time-to-death of 351 HIV-positive adults on antiretroviral therapy (ART) from 2015 to 2019. It used criteria like AIC and BIC,checking from all the models used the Weibull-Gamma shared frailty model was found as the model that had the best fit, which helped in accounting for significant heterogeneity between patient residences. The Key indicators and predictors of survival included age, marital status, education level, TB status, opportunistic infections, WHO clinical stage, CD4 count, and weight. The study also highlighted high rates of mortality that occurred within the first 10 months of ART initiation, This underscored the need of addressing patient outcomes variability to help in improving effectiveness in ART. Frailty terms like Kendall's tau confirmed the dependence within clusters, emphasizing the need for frailty models in understanding disparities in survival outcomes among HIV/AIDS patients (Khan and Awan, 2017).

Frailty models have also been applied effectively in studies on child mortality to help in accounting for unobserved heterogeneity at the maternal and community levels. In Bangladesh, significant progress in reduction of under-five mortality has been noted, yet there still exist disparities, particularly in the Sylhet division, which experiences higher child mortality due to factors such as lower healthcare access and cultural practices. Cox proportional hazards models with frailty components revealed that maternal age, education, birth order, birth interval, and maternal employment are important child mortality predictors. The frailty effects highlight unmeasured factors like genetic predispositions, parental competence, and healthcare quality. This emphasizes the need for comprehensive public health interventions focusing on maternal education, birth spacing, and reducing early motherhood, as well as addressing community-level determinants to achieve sustainable reductions in child mortality (Belay and Derebe, 2022).

The application of frailty models, particularly the spatially correlated Conditional Autoregressive (CAR) frailty model, has provided significant insights into malaria transmission dynamics, especially in areas impacted by environmental changes such as dam construction. A study conducted in southwestern Ethiopia examined the effects of the Gilgel-Gibe hydroelectric dam on malaria incidence among children under 10. The CAR frailty model accounted for spatial clustering, showing that malaria risk decreased with distance from the dam (HR=0.95 per km) (Wondaya et al., 2016).

2.8 Latent factors In malaria-related health behaviors and access to preventive care

Malaria remains a public health challenge that is very crucial, to be more specific in sub-Saharan Africa, where it affects vulnerable populations disproportionately. Effective prevention and treatment rely not only on the availability of interventions such as insecticide-treated nets (ITNs) and antimalarial drugs but also on understanding the latent factors influencing health behaviors and access to care. These latent factors, include determinants that unob-

servable or indirect which include socioeconomic status, cultural norms, and individual perceptions, that play a crucial role in the shaping of malaria prevention and practices in treatment.

In a study conducted in Ethiopia it highlights the critical role of psychosocial and contextual factors in shaping malaria-related health behaviors, most particularly in care-seeking for febrile children under five in rural Ethiopia. Prompt care-seeking, clearly elaborated as seeking treatment and care within 24 hours of fever onset, was contributed by factors that are ideational such as caregivers' self-efficacy, response efficacy, attitudes toward prompt care-seeking, involvement in household decision-making, and perceptions of gender equity. Contextual factors, including region of residence, caregiver education level, and household vulnerability, also played significant roles. These findings align with broader research across Ethiopia and sub-Saharan Africa, which shows that higher education levels, household income, and positive perceptions of antimalarial efficacy improve care-seeking behaviors. The study underscores the need for multi-sectoral interventions targeting both individual ideation and structural barriers, emphasizing community-level social behavior change and health system improvements to enhance malaria prevention and timely treatment (Wondaya et al., 2016). Socioeconomic determinants have a major role in helping in malaria prevention behaviors, as evidenced by study in the North West Region of Cameroon. Key factors influencing households' adoption of preventive measures include knowledge of malaria signs and causes, community malaria prevalence, household income, size, education level, and employment status of household heads. Age and marital status of household heads further influence behavior, highlighting the intersection of demographic and socioeconomic variables. Despite the availability of basic prevention methods, gaps in knowledge about malaria signs persist, emphasizing the need for targeted sensitization campaigns and public health awareness. Recommended interventions include the formation of community-based malaria control committees, subsidized education to enhance awareness, and the free distribution of preventive tools like ITNs and insecticides. The study underscores the importance of communal rather than individualistic strategies, suggesting that collective efforts yield more sustainable malaria prevention outcomes, particularly in low-resource settings. The study shows there is still knowledge gap on this issue which necessitates actions towards public

health awareness in relation to malaria signs which would be a cause to really know the variability caused by awareness (Yakum et al., 2020). The study on school-based malaria control in Kenya (Halliday, 2015) evaluated the impact of intermittent screening and treatment (IST) for malaria among schoolchildren. Conducted in the coastal districts of Kwale and Msambweni, the trial aimed to assess how IST affects children's health, education, and sustained attention. Findings indicated that while the intervention reduced malaria-related health risks, its impact on educational outcomes was minimal. The study concluded that a targeted, integrated approach combining health and education interventions is essential to address the heterogeneous risks and challenges faced by schoolchildren in malaria-endemic regions.Hence these gives a need to look at the latent child specific factors in regards to Malaria-related health behaviors and access to preventive care .

The study in rural Cameroon found that the proper use of insecticide-treated nets (ITNs) significantly reduced malaria prevalence in children under five. Higher education levels correlated with better use of antimalarial drugs, and treatment outcomes varied based on healthcare preferences, particularly quinine use. Socioeconomic status also influenced health-seeking behavior, with wealthier households more likely to seek proper malaria care. The study underscores the importance of malaria education and integrated prevention, advocating for partnerships with local communities and organizations like the Global Fund to enhance resource access and malaria control efforts.Hence in underscoring on the need of malaria education hence the need to have more investigation on digital health .Since the study mostly focused on rural area hence a need to have more representative sample and also need to focus on endemic and non-endemic regions (Azunie, 2017).

2.9 Conclusions

Despite progress in malaria control, significant gaps remain in understanding how geographic, socioeconomic, and child-specific factors influence prevention behaviors, especially in rural areas. Child-specific factors, such as age, adherence to antimalarial treatments, and vulnerability to symptoms, are underexplored. Additionally, the role of digital connectivity in

addressing disparities is limited. This study uses inverse Gaussian frailty models to account for unobserved heterogeneity in health behaviors, effectively capturing the complex, clustered nature of malaria-related data and offering insights for developing equitable, child-centered malaria prevention strategies.

2.10 Current research

This dissertation, has The primary aim of using inverse Gaussian frailty models in this context to capture and quantify unobserved heterogeneity in malaria prevention behaviors, accounting for latent factors at individual, household, and community levels. Specifically, the model seeks to analyze how geographic, socioeconomic, and child-specific factors like age and treatment adherence contribute to malaria risks. Achieving these objectives will provide deeper insights into the clustering of malaria outcomes, enabling the design of tailored, equitable interventions that target high-risk populations, particularly vulnerable children in rural areas.

Chapter 3

Methodology

3.1 Introduction

This chapter provides a comprehensive and detailed explanation of the methodology adopted to achieve the objectives of this study. The focus is on utilizing semi-parametric inverse Gaussian frailty models to analyze unobserved heterogeneity in malaria prevention behaviors and access to care. The chapter is organized to discuss the research design, data sources, and variable selection , followed by a detailed description of the statistical models employed, including their formulation and parameter estimation. A step-by-step analytical approach is presented to ensure clarity in the application of the methods.

3.2 Data

This study utilized secondary data from the 2020 KDHS, a nationally representative survey conducted by the Kenya National Bureau of Statistics (KNBS) in collaboration with other stakeholders. The KDHS data was collected between November 9 and December 19, 2020, employing a stratified two-stage cluster sampling design to ensure accurate representation of households and individuals across Kenya. The survey captures comprehensive health-related information, including household characteristics, socioeconomic status, and individual health indicators. Data collection involved face-to-face interviews using structured questionnaires administered to eligible respondents.

3.3 Data preprocessing

Before conducting the exploratory data analysis, the dataset was assessed for missing values. The missing data percentages for key variables were as follows in Table 3.1:

Table 3.1: Missing data summary

Variable	n_miss	pct_miss
act_adherence	2881	79.5
sm_phone_use	793	21.9
treatednet	190	5.24
child_age	45	1.24
malaria_symptom	14	0.386
hold_cluster	0	0
endemic_zones	0	0
internet_use	0	0
socialmediause	0	0
time_to_itnusage	0	0

The *MCAR (Missing completely at random) test* that evaluates whether the missing data in a dataset occur purely at random, meaning that the probability of missingness is unrelated to both observed and unobserved data. In this case, the test was conducted using `mcar_test(new_data2)`, yielding a test statistic of **180**, with **27 degrees of freedom (df)** and a **p-value of 0**. Since the p-value is extremely small ($p < 0.05$), we reject the null hypothesis that the data are missing completely at random. This suggested that the missingness in the dataset followed a systematic pattern rather than occurring randomly. Additionally, the output indicated that there are **11 distinct missing data patterns**, implying potential dependencies between missing values and observed variables. Given this result, further investigation into *Missing at Random (MAR) or Missing Not at Random (MNAR)* mechanisms was necessary, and appropriate imputation methods such as *Multiple imputation by chained equations (MICE)* should be considered to handle the missing data.

The logistic regression results in Table 3.2 indicated that missingness depended on observed variables:

Table 3.2: Logistic regression results for missingness

Variable	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.84122	0.17041	-4.936	7.95e-07
child_age	-0.15911	0.06792	-2.343	0.01915
trtdnet_std	-0.41749	0.17757	-2.351	0.01872
act_adherence_std	0.58577	0.20288	2.887	0.00389

Child age and ITN usage were associated with lower odds of missingness, reducing the likelihood by 15% and 34%, respectively. In contrast, ACT adherence was linked to higher odds of missingness, with an observed 80% increase. These findings support the **Missing at random (MAR)** assumption, where missingness depends on observed data but not on unobserved data. **Predictive mean matching (PMM)** was chosen as the imputation method in **Multiple imputation by chained equations (MICE)** because it is well-suited for handling missing continuous data while preserving the underlying distribution. PMM imputes missing values by selecting observed values from the most similar cases based on predicted means, reducing the risk of unrealistic imputations. Additionally, performing multiple imputations ($m = 5$) helps account for uncertainty in the missing data, leading to more robust statistical inferences.

3.4 Data analysis

The data used in this study comprised of observations related to malaria prevention and treatment behaviors, structured within a cross-sectional framework. The underlying processes in the data may reflect long-term patterns and variability influenced by household (socioeconomic status, household size, digital connectivity), individual (child age, malaria symptoms, ACT adherence), and geographic factors data (endemic vs. non-endemic zones).

These latent structures can reveal unobserved heterogeneity at different levels, such as household or cluster-level effects, which impact malaria-related behaviors. To capture these complexities, the study employed semi-parametric inverse Gaussian frailty models. The inverse Gaussian frailty was selected over gamma frailty due to its superior performance in scenarios with extreme survival times or highly variable cluster-specific effects, as evidenced by its heavier-tailed distribution. Compared to the standard Cox model, it explicitly accounts for within-cluster dependency a critical advantage given the hierarchical structure of our malaria data (households nested within geographic clusters). This specification outperforms alternatives like the gamma frailty model when the unobserved heterogeneity exhibits heavier tails or skewness, and it is more flexible than the standard Cox model in accounting for cluster-specific random effects. Based on the reviewed literature and the data's hierarchical nature, this modeling approach will be particularly suitable for quantifying unobserved variability and its influence on malaria prevention and treatment behaviors, offering insight into likely outcomes under similar conditions.

3.5 Model formulation

3.5.1 Formulation of inverse Gaussian frailty model

The inverse Gaussian frailty model is a statistical approach used to analyze time-to-event data, accounting for unobserved heterogeneity among clusters or individuals. Frailty models introduced a random effect, or "frailty," to capture latent variables that influence the hazard function. In the inverse Gaussian frailty model, the frailty term was assumed to follow an inverse Gaussian distribution, which was well-suited for modeling skewed random effects. The use of semi-parametric inverse Gaussian frailty models in this study is guided by their unique ability to account for unobserved heterogeneity in clustered data, which is a critical feature of malaria prevention and treatment behaviors. Additionally, the semi-parametric nature of the model will provide the advantage of not requiring a strict parametric form for

the baseline hazard, making it ideal for analyzing complex datasets like the KDHS, where time-to-event data is proxied by health behavior outcomes. The following assumptions apply:

i) **Proportional hazards:** The hazard function is multiplicative,it will be assumed that covariates have a proportional effect on the baseline hazard. This was verified using the Schoenfeld residuals where if the pvalue <0.05 then assumption is violated. The plots in Figure 3.1 and corresponding statistical results from the Cox proportional hazards model in Table 3.3 indicate that the proportional hazards assumption holds for the covariates, as evidenced by the relatively stable beta estimates across time for `endemic_std`, `soc_media_std`, `int_use_std`, and `mal_symptoms_std`, with p-values indicating non-significance (all > 0.05).

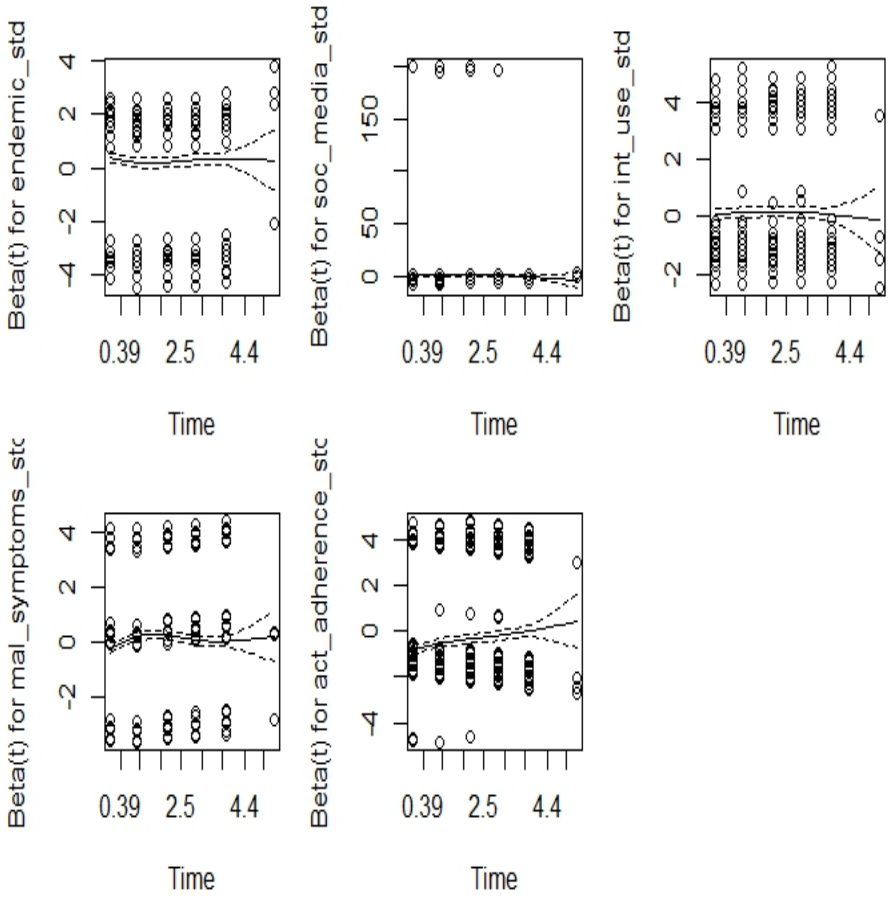


Figure 3.1: Diagnostic plots assessing the proportional hazards assumption for various covariates in the Cox model.

Table 3.3: Proportional Hazards Assumption Tests

Covariate	χ^2	**df**	**p-value**
endemic_std	0.2928	1	0.5884
soc_media_std	0.0652	1	0.7985
int_use_std	0.8354	1	0.3607
mal_symptoms_std	1.6655	1	0.1969

ii) **Frailty distribution:** The frailty term u_j will follow an inverse Gaussian distribution with mean 1 and variance θ, capturing variability across clusters.

iii) **Independence of frailties:** frailty terms u_j will be assumed to be independent across clusters, ensuring no correlation between unobserved effects of different clusters. After fitting the frailty model, the residuals will be analyzed for potential correlation between clusters. If the frailties are truly independent, no significant correlation should be detected.

iv) **Covariates Independence:** The covariates x_{ij} will be assumed to be independent of the frailty term u_j.After estimating the frailty term (u_j), correlations between u_j and the covariates ($x_i j$) will be computed.

The inverse Gaussian frailty model was justified in this dissertation as it models the frailty term u_j as an inverse Gaussian random variable to account for unobserved heterogeneity, which is critical in clustered data where shared latent characteristics, such as socioeconomic status, access to healthcare, or community-level behaviors, can influence the risk rate. The inverse Gaussian distribution is particularly well suited for capturing skewed and heavy-tailed random effects, making it ideal for modeling heterogeneity in health behaviors or disease risk, especially when the variance of frailty across clusters is large and asymmetrically distributed.

This approach provides a flexible framework by incorporating both fixed covariates (x_{ij}) and random effects (u_j), allowing a comprehensive assessment of the interaction between observed and unobserved influences. Compared to gamma frailty models, the inverse

Gaussian frailty model tends to perform better when the frailty distribution is highly skewed or when extreme cluster-level variation exists. Although the gamma distribution assumes a relatively moderate level of heterogeneity, the inverse Gaussian can capture more substantial deviations and long-tail behavior, making it more robust in settings where a few clusters exhibit markedly different risks due to latent factors.

Moreover, the shared frailty structure of the inverse Gaussian model offers advantages over the standard Cox proportional hazards model by effectively addressing intra-cluster correlation and dependence among observations. In contrast, the Cox model assumes independence among observations and lacks the capacity to model latent heterogeneity. Thus, the inverse Gaussian frailty model outperforms both the gamma frailty and Cox models in scenarios characterized by strong unobserved heterogeneity, significant right-skewness in frailty effects, and greater variability across clusters, conditions often observed in population-based studies of infectious diseases or health behaviors influenced by socioeconomic disparities.

The parameters of the inverse Gaussian frailty model, including the regression coefficients (β), baseline hazard ($\lambda_0(t)$), and frailty variance (θ), are estimated using maximum likelihood estimation (MLE), a method that estimates parameters by maximizing the likelihood function given the observed data. MLE is chosen for its desirable statistical properties: It ensures consistency, as the estimates converge to the true parameter values with increasing sample size; it offers efficiency by achieving the lowest possible variance among unbiased estimators, yielding precise estimates; and it provides flexibility to handle complex likelihoods involving frailty terms and random effects. Additionally, MLE enables robust statistical inference through hypothesis testing and confidence interval estimation by leveraging the asymptotic normality of the estimators, while also delivering unbiased estimates in larger samples, making it a reliable approach for parameter estimation in this modeling framework.

$$S_{ij}(t \mid u_j) = \exp\left(-u_j \int_0^t \lambda_0(s) e^{\beta^T x_{ij}}\, ds\right), \tag{3.1}$$

In Equation (3.1) the survival function with frailty, starts by considering the survival probability for an individual i in cluster j, accounting for unobserved heterogeneity captured through the frailty term u_j. Where:

i) t: Time variable, that represents time to usage of insecticide treated nets .

ii) $\lambda_0(t)$: Baseline hazard function, representing the hazard rate if all covariates are zero.

iii) x_{ij}: Covariate vector for individual i in cluster j.

iv) β: Vector of regression coefficients, quantifying the effect of covariates on the hazard.

v) u_j: Cluster-specific frailty term, which will follow an inverse gaussian distribution, which introduces random effects to model unobserved heterogeneity.

$$\lambda_0(t) = \lambda e^{\gamma t}, \tag{3.2}$$

The gompertz distribution is characterized by a baseline hazard function of the form as given in Equation (3.2). where:

i) $\lambda > 0$ is the scale parameter (baseline hazard at $t = 0$),

ii) γ is the shape parameter (determines whether the hazard increases or decreases over time).

$$S_{ij}(t \mid u_j) = \exp\left(-u_j \int_0^t \lambda e^{\gamma s} e^{\beta^T x_{ij}}\, ds\right). \tag{3.3}$$

The survival function derivation employs the gompertz baseline hazard form $\lambda_0(t) = \lambda e^{\gamma t}$, with the complete substitution given in Equation (3.3).

After simplifying the integral we get the survival function as Equation (3.7) :

$$\int_0^t \lambda e^{\gamma s} e^{\beta^T x_{ij}}\, ds = \lambda e^{\beta^T x_{ij}} \int_0^t e^{\gamma s}\, ds. \tag{3.4}$$

The integral $\int_0^t e^{\gamma s}\,ds$ can be solved as:

$$\int_0^t e^{\gamma s}\,ds = \frac{e^{\gamma t}-1}{\gamma}. \tag{3.5}$$

The survival function can be expressed as

$$S_{ij}(t \mid u_j) = \exp\left(-u_j \lambda e^{\beta^T x_{ij}} \cdot \frac{e^{\gamma t}-1}{\gamma}\right). \tag{3.6}$$

$$S_{ij}(t \mid u_j) = \exp\left(-\frac{u_j \lambda e^{\beta^T x_{ij}}}{\gamma}\left(e^{\gamma t}-1\right)\right). \tag{3.7}$$

The survival function with the gompertz baseline hazard is given by Equation (3.7).

This formulation emphasizes that shared frailty u_j will affect survival probabilities within clusters. The model will be particularly useful for studying household or regional effects on health outcomes in cross-sectional data, where time may be treated as constant.

$$\lambda_{ij}(t \mid u_j) = u_j \lambda_0(t) e^{\beta^T x_{ij}}, \tag{3.8}$$

The hazard function as described by Adham and AlAhmadi (2016) for the i-th individual in the j-th cluster is given by Equation (3.8).Where:

i) $\lambda_0(t)$: The baseline hazard function.

ii) $\mathbf{X}_{ij}$: A vector of covariates for the i-th individual in the j-th cluster.

iii) β: The regression coefficients.

iv) u_j: The frailty term specific to the j-th cluster, following an inverse gaussian distribution with mean 1 and variance θ.

It helps to model the instantaneous risk that the event occurs at time t given the frailty term u_j for the cluster.

In this case The gompertz distribution was chosen for the parametric frailty model due to its flexibility in modeling hazard rates and its successful convergence with the data. Other distributions (example: exponential, Weibull, loglogistic, lognormal) failed to converge, likely due to a mismatch between their assumptions and the data's hazard rate shape. The gompertz distribution's ability to model both increasing and decreasing hazard rates makes it well-suited for analyzing time-to-event data in this context.

To replace the baseline hazard $\lambda_0(t)$ in the equation with the gompertz baseline hazard, we substitute:

$$\lambda_0(t) = \lambda e^{\gamma t}, \tag{3.9}$$

where:

i) $\lambda > 0$ is the scale parameter,

ii) γ is the shape parameter.

The updated equation becomes:

$$\lambda_{ij}(t \mid u_j) = u_j \lambda e^{\gamma t} e^{\beta^T x_{ij}}. \tag{3.10}$$

$$\lambda_{ij}(t \mid u_j) = u_j \lambda e^{\gamma t + \beta^T x_{ij}}. \tag{3.11}$$

The final simplified form of the equation of the hazard function is given by Equation (3.11)

$$f(u_j; \mu, \theta) = \frac{1}{\sqrt{2\pi\theta u_j^3}} \exp\left(-\frac{(u_j - \mu)^2}{2\theta u_j}\right) \tag{3.12}$$

The frailty u_j is modeled using the inverse gaussian distribution as given by Equation (3.12)

for

$$\mu_j > 0, \quad \text{where } \mu > 0 \text{ is the mean, and } \theta > 0 \text{ is the shape parameter.}$$

where :

i) μ: Mean of the frailty term (often set to 1 for identifiability).

ii) θ: Variance of the frailty term, will represent unobserved heterogeneity.

The joint likelihood will integrate over the frailty term, ensuring it accounts for random effects shared within clusters. For individual i in cluster j the likelihood involves the conditional density of the outcome given frailty and the frailty distribution itself. The combined likelihood for multiple clusters will multiply the likelihoods of the individual clusters.The likelihood will be given by first obtaining the marginal survival function which is given by:

$$L_j = \int_0^{\infty} \prod_{i=1}^{n_j} f(T_{ij} \mid u_j) f(u_j)\, du_j \tag{3.13}$$

The likelihood for an individual i in cluster j is given by Equation (3.13) , where:

i) $f(T_{ij} \mid u_j)$ is the conditional density of the survival time T_{ij} given the frailty u_j.

ii) $f(u_j)$ is the probability density function (PDF) of the frailty term u_j, which follows an inverse gaussian distribution.

$$f(T_{ij} \mid u_j) = \lambda_{ij}(T_{ij} \mid u_j)^{\delta_{ij}} \cdot S_{ij}(T_{ij} \mid u_j) \tag{3.14}$$

The conditional density $f(T_{ij} \mid u_j)$ can be expressed in terms of the hazard function $\lambda_{ij}(t \mid u_j)$ and the survival function $S_{ij}(t \mid u_j)$ as given by Equation (3.14) , where:

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

ii) $\lambda_{ij}(T_{ij} \mid u_j)$ is the hazard function for individual i in cluster j.

iii) $S_{ij}(T_{ij} \mid u_j)$ is the survival function for individual i in cluster j.

From the model, we have the hazard function given by :

$$\lambda_{ij}(t \mid u_j) = u_j \eta e^{\gamma t} e^{\beta^T x_{ij}} \tag{3.15}$$

where:

i. $\lambda_{ij}(t \mid u_j)$: Hazard function for individual i in cluster j at time t, conditional on frailty.

ii. u_j: Cluster-specific frailty term (assumed to follow an inverse Gaussian distribution with mean 1 and variance θ); captures unobserved heterogeneity shared within the cluster.

iii. η: Baseline hazard scaling parameter; adjusts the overall level of the baseline hazard.

iv. $e^{\gamma t}$: Time-dependent exponential term.

$\gamma > 0$: hazard increases over time.

$\gamma < 0$: hazard decreases over time.

v. x_{ij}: Covariate vector for individual i in cluster j (example: age, sex, education, behaviors).

vi. β: Vector of regression coefficients; represents the effect size and direction of each covariate.

vii. $e^{\beta^T x_{ij}}$: Relative risk due to the covariate profile of individual i in cluster j.

We have the survival function given by :

$$S_{ij}(t \mid u_j) = \exp\left(-u_j \eta e^{\beta^T x_{ij}} \frac{e^{\gamma t} - 1}{\gamma}\right) \tag{3.16}$$

Substituting these into the expression for $f(T_{ij} \mid u_j)$ we have :

$$f(T_{ij} \mid u_j) = \left(u_j \eta e^{\gamma T_{ij}} e^{\beta^T x_{ij}}\right)^{\delta_{ij}} \cdot \exp\left(-u_j \eta e^{\beta^T x_{ij}} \frac{e^{\gamma T_{ij}} - 1}{\gamma}\right) \tag{3.17}$$

$$f(u_j) = \frac{1}{\sqrt{2\pi\theta u_j^3}} \exp\left(-\frac{(u_j - \mu)^2}{2\theta u_j}\right) \tag{3.18}$$

Incorporating the frailty distribution we have the frailty term u_j which follows an Inverse Gaussian distribution with PDF: The probability density function of the frailty term u_j is given by Equation (3.18),

where $\mu > 0$ is the mean and $\theta > 0$ is the variance parameter.

Substituting $f(T_{ij} \mid u_j)$ and $f(u_j)$ into the likelihood function L_j of Equation (3.13) we have :

$$L_j = \int_0^\infty \prod_{i=1}^{n_j} \left[\left(u_j \eta e^{\gamma T_{ij}} e^{\beta^T x_{ij}} \right)^{\delta_{ij}} \cdot \exp\left(-u_j \eta e^{\beta^T x_{ij}} \frac{e^{\gamma T_{ij}} - 1}{\gamma} \right) \right] \cdot f(u_j)\, du_j \tag{3.19}$$

Simplifying the product over individuals $i = 1, \ldots, n_j$:

$$L_j = \int_0^\infty \left(\prod_{i=1}^{n_j} \left[u_j \eta e^{\gamma T_{ij}} e^{\beta^T x_{ij}} \right]^{\delta_{ij}} \right) \cdot \exp\left(-u_j \sum_{i=1}^{n_j} \eta e^{\beta^T x_{ij}} \frac{e^{\gamma T_{ij}} - 1}{\gamma} \right) \cdot f(u_j)\, du_j \tag{3.20}$$

The final likelihood function incorporates:

i. The product of the hazard terms for individuals who experience the event ($\delta_{ij} = 1$):

$$\prod_{i=1}^{n_j} \left[u_j \eta e^{\gamma T_{ij}} e^{\beta^T x_{ij}} \right]^{\delta_{ij}} \tag{3.21}$$

ii. The survival terms for all individuals:

$$\exp\left(-u_j \sum_{i=1}^{n_j} \eta e^{\beta^T x_{ij}} \frac{e^{\gamma T_{ij}} - 1}{\gamma} \right) \tag{3.22}$$

iii. The frailty distribution $f(u_j)$:

$$\frac{1}{\sqrt{2\pi\theta u_j^3}} \exp\left(-\frac{(u_j - \mu)^2}{2\theta u_j} \right) \tag{3.23}$$

$$\begin{aligned} L_j = \int_0^\infty & \left(\prod_{i=1}^{n_j} \left[u_j \eta e^{\gamma T_{ij}} e^{\beta^T x_{ij}} \right]^{\delta_{ij}} \right) \cdot \exp\left(-u_j \sum_{i=1}^{n_j} \eta e^{\beta^T x_{ij}} \frac{e^{\gamma T_{ij}} - 1}{\gamma} \right) \\ & \cdot \frac{1}{\sqrt{2\pi\theta u_j^3}} \exp\left(-\frac{(u_j - \mu)^2}{2\theta u_j} \right) du_j \end{aligned} \tag{3.24}$$

Thus, the final likelihood function is given by Equation (3.24)

Substituting $f(T_{ij} \mid u_j)$ using the new hazard function and survival function, the likelihood function will incorporate:

$$\prod_{i=1}^{n_j} \left[u_j \lambda e^{\gamma T_{ij}} e^{\beta^\top x_{ij}} \right]^{\delta_{ij}} \tag{3.25}$$

where:

i. δ_{ij} is an event indicator (1 if the event occurs, 0 otherwise).

ii. The integral over u_j accounts for the frailty.

The log-likelihood function will be easier to maximize hence taking the natural logarithm of the likelihood function we have :

$$\ell(\mu, \theta \mid u_1, \ldots, u_n) = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \delta_{ij} \log \left(u_j \lambda e^{\gamma T_{ij}} e^{\beta^\top x_{ij}} \right) + \log f(u_j) \tag{3.26}$$

Breaking it down:

$$\ell(\mu, \theta) = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \delta_{ij} \left(\log u_j + \log \lambda + \gamma T_{ij} + \beta^\top x_{ij} \right) + \sum_{j=1}^{m} \log f(u_j) \tag{3.27}$$

To estimate the parameters μ and θ using MLE, we take partial derivatives of the log-likelihood function with respect to μ and θ , set them equal to zero, and solve for the parameters. Derivative with respect to μ:

$$L_{ij} = \left[u_j \lambda e^{\gamma T_{ij}} e^{\beta^T x_{ij}} \right]^{\delta_{ij}} \exp \left(-u_j \frac{\lambda}{\gamma} \left(e^{\gamma T_{ij}} - 1 \right) e^{\beta^T x_{ij}} \right). \tag{3.28}$$

The likelihood contribution for individual i in cluster j is given by Equation (3.28):

$$L_j = \int_0^\infty \left[\prod_{i=1}^{n_j} \left(u_j \lambda e^{\gamma T_{ij}} e^{\beta^T x_{ij}} \right)^{\delta_{ij}} \exp \left(-u_j \sum_{i=1}^{n_j} \frac{\lambda}{\gamma} \left(e^{\gamma T_{ij}} - 1 \right) e^{\beta^T x_{ij}} \right) \right] f(u_j)\, du_j. \tag{3.29}$$

Likelihood for allindividuals in cluster j is given by Equation (3.29)

Substituting the inverse gaussian frailty distribution:

$$f(u_j) = \frac{1}{\sqrt{2\pi\theta u_j^3}} \exp\left(-\frac{(u_j-\mu)^2}{2\theta u_j}\right). \tag{3.30}$$

The total likelihood function across clusters is given by :

$$L = \prod_{j=1}^{m} L_j. \tag{3.31}$$

Taking the log-likelihood:

$$\ell(\mu,\theta,\lambda,\gamma,\beta) = \sum_{j=1}^{m}\left[\sum_{i=1}^{n_j} \delta_{ij} \log\left(u_j \lambda e^{\gamma T_{ij}} e^{\beta^T x_{ij}}\right) + \log f(u_j)\right]. \tag{3.32}$$

$$\ell(\mu,\theta) = \sum_{j=1}^{m}\sum_{i=1}^{n_j} \delta_{ij}\left(\log u_j + \log\lambda + \gamma T_{ij} + \beta^T x_{ij}\right) + \sum_{j=1}^{m} \log f(u_j). \tag{3.33}$$

In Equation (3.33) the log-likelihood function was expanded.

For maximum likelihood estimation we get the partial derivatives

$$\frac{\partial \ell}{\partial \mu} = \sum_{j=1}^{m}\sum_{i=1}^{n_j} \frac{u_j - \mu}{\theta u_j} = 0. \tag{3.34}$$

To estimate μ, the mean of the frailty distribution, we compute the derivative with respect to μ as given in Equation (3.34).

$$\mu = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j} u_j}{n}. \tag{3.35}$$

The maximum likelihood estimate of μ is obtained by solving the score equation, resulting in Equation (3.35).

$$\frac{\partial \ell}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j} \frac{(u_j-\mu)^2}{u_j} = 0. \tag{3.36}$$

The estimate for θ, the variance of the frailty distribution, is obtained by solving the derivative with respect to θ in Equation (3.36).

$$\theta = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n_j} \frac{(u_j - \mu)^2}{u_j}. \tag{3.37}$$

The maximum likelihood estimate of θ is obtained by solving the score equation, resulting in Equation (3.37).

$$\frac{\partial \ell}{\partial \lambda} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(\frac{\delta_{ij}}{\lambda} - \frac{u_j}{\gamma} \left(e^{\gamma T_{ij}} - 1 \right) e^{\beta^T x_{ij}} \right) = 0. \tag{3.38}$$

The estimate for λ, the scale parameter of the baseline hazard, is obtained by taking the derivative with respect to λ in Equation (3.38).

$$\lambda = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} \delta_{ij}}{\sum_{j=1}^{m} \sum_{i=1}^{n_j} \frac{u_j}{\gamma} \left(e^{\gamma T_{ij}} - 1 \right) e^{\beta^T x_{ij}}}. \tag{3.39}$$

The maximum likelihood estimate of λ is obtained by solving the score equation, resulting in Equation (3.39).

$$\frac{\partial \ell}{\partial \gamma} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(\delta_{ij} T_{ij} - \frac{u_j \lambda e^{\beta^T x_{ij}}}{\gamma^2} e^{\gamma T_{ij}} T_{ij} \right) = 0. \tag{3.40}$$

The estimate for γ, the shape parameter of the gompertz distribution, is obtained by differentiating with respect to γ in Equation (3.40).

Solving for γ involves numerical methods.

$$\frac{\partial \ell}{\partial \beta} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(\delta_{ij} x_{ij} - \frac{u_j \lambda}{\gamma} \left(e^{\gamma T_{ij}} - 1 \right) e^{\beta^T x_{ij}} x_{ij} \right) = 0. \tag{3.41}$$

The estimates for β, the covariate effects, are obtained by taking the derivative with respect to β in Equation (3.41).Solving for β requires an iterative approach (example:Newton-Raphson).

The variability in frailty across households or clusters is measured by the parameter θ, which quantifies the extent of unobserved heterogeneity.The parameter θ is estimated to assess the degree of heterogeneity in outcomes between clusters. A large value of θ that significantly differs from zero indicates a substantial variability between clusters, which reflects the presence of unobserved factors contributing to differences in the outcome. Additionally, a large θ signifies a strong association among individuals within the same cluster, suggesting shared influences or dependencies. Conversely, when θ=0, it implies that the frailties are uniformly equal to one, indicating an absence of cluster-specific effects. In this case, outcomes are independent both within and across clusters, with no evidence of unobserved heterogeneity.

Covariate Effects (β): they help Interpret the estimated regression coefficients to understand the impact of each covariate on the log-relative risk of the event.Positive events show an increase in the likelihood in the ITN usage when event =1. Negative values of β show a decrease in the likelihood of usage.

In distinguishing the Gamma frailty model, The frailty term u_i is assumed to follow a Gamma distribution with shape parameter $\frac{1}{\theta}$ and rate parameter $\frac{1}{\theta}$, which implies a mean of 1 and a variance of θ.:

$$u_i \sim \text{Gamma}\left(\frac{1}{\theta}, \frac{1}{\theta}\right) \tag{3.42}$$

This distribution is strictly positive, right-skewed, and has a closed-form expression for the marginal survival function, making it computationally convenient. The marginal survival function for an individual j in cluster i, with covariates X_{ij}, is:

$$S_{ij}(t) = \left(1 + \theta H_0(t) \exp(X_{ij}^{\top} \beta)\right)^{-1/\theta} \tag{3.43}$$

where $H_0(t)$ is the baseline cumulative hazard function, and β is the vector of regression coefficients.

In contrast, the Inverse Gaussian frailty model assumes that the frailty term u_i follows an Inverse Gaussian distribution, also with mean 1 and variance θ:

$$u_i \sim \text{IG}(1, \theta) \tag{3.44}$$

The Inverse Gaussian distribution has a heavier tail than the Gamma distribution, which makes it better suited for data with more extreme frailty values or clusters exhibiting more pronounced unobserved heterogeneity. Although it does not yield a simple closed-form expression for the marginal survival function like the Gamma model, it provides more flexibility in modeling long-tailed unobserved effects. The marginal survival function in this case involves the Laplace transform of the Inverse Gaussian distribution:

$$S_{ij}(t) = \exp\left\{\frac{1}{\theta}\left(1 - \sqrt{1 + 2\theta H_0(t)\exp(X_{ij}^\top \beta)}\right)\right\} \tag{3.45}$$

The choice between these two models depends on the distribution of the unobserved heterogeneity. The Gamma frailty model is simpler and often sufficient for moderate levels of heterogeneity, while the Inverse Gaussian frailty model is more appropriate when the heterogeneity is more extreme or when heavier-tailed distributions are observed in the data. The Inverse Gaussian model better accommodates clusters with more substantial deviation from the average risk due to its heavier tail.

3.5.2 Model evaluation

Deviance residuals will assess how well a survival model aligns with observed data by quantifying the difference between observed and expected events for each individual. Residuals will be plotted against fitted values to detect systematic deviations or against covariates to uncover relationships not captured by the model. Large residuals will indicate potential outliers, while systematic patterns will suggest model misfit, missing covariates, or violated assumptions, guiding improvements to model accuracy and fit.

Baseline hazard inspection will assess whether ($\lambda_0(t)$) accurately represents the underlying risk when covariates are zero. For time-to-event data, the baseline or cumulative hazard will be extracted and plotted to identify expected trends, such as increasing or decreasing risk over time. For binary outcomes (t=1), the baseline hazard is treated as a constant representing the event rate for individuals with average covariates. A smooth hazard curve will indicate a good fit, while unexpected patterns or deviations will suggest potential model misspecifications, unaccounted heterogeneity, or the need for stratified or time-varying covariates.

3.5.3 Ethical consideration

The study adhered to ethical guidelines for secondary data analysis, ensuring confidentiality and proper use of KDHS data. No identifiable information was included in the analysis. Ethical approval for the KDHS was granted by the Kenyan National Ethics Review Committee and the Institutional Review Board of the implementing agency. Informed consent was obtained from all participants, and no personally identifiable information was retained. Data were anonymized at the cluster level, with household identifiers replaced by randomly generated codes. Use of the KDHS data is governed by a Data Use Agreement with the DHS Program, which prohibits re-identification or unauthorized sharing. Compliance is enforced through audits and potential revocation of access, ensuring strict confidentiality and ethical standards throughout the study.

3.5.4 Conclusion

This methodology provided a robust framework for evaluating ITN usage in malaria-endemic regions, leveraging frailty models with inverse Gaussian distributions to account for both individual-level and cluster-level unobserved heterogeneity. By integrating predictors such as geographic, socioeconomic, and child-specific factors, the model captured the complex dynamics influencing ITN adoption behaviors. The semi-parametric approach ensured flexibility in modeling baseline hazards, while the frailty terms highlighted latent variability across clusters, enabling a deeper understanding of contextual influences. This approach

yielded more precise and reliable estimates, offering valuable insights into health behavior patterns and guiding targeted malaria prevention interventions.

Chapter 4

Results and interpretation

4.1 Introduction

This chapter presents the results of the exploratory data analysis, diagnostic tests, and findings from the fitted *Inverse Gaussian Frailty* models to assess the impact of malaria-endemic zones, digital connectivity, and child health factors on malaria-related health behaviors.

4.2 Descriptive statistics

The Table 4.1 summarizes the number of households and the percentage of ITN usage in malaria-endemic and non-endemic zones. The results suggest that ITN usage is higher in endemic zones compared to non-endemic zones, which may reflect greater awareness or accessibility of preventive measures in areas with higher malaria risk. A chi-square test showed a significant association between ITN usage and endemicity ($\chi^2 = 23.9$, $p < 0.001$), indicating that households in endemic zones are more likely to use treated nets.

Table 4.1: Household itn usage by endemic zone

Endemic Zone	**Households (n)**	**ITN Use (%)**
Endemic	2381	54.2
Non-Endemic	1055	45.1

The analysis as per Table 4.2 reveals significant associations between ITN usage and two digital connectivity variables:

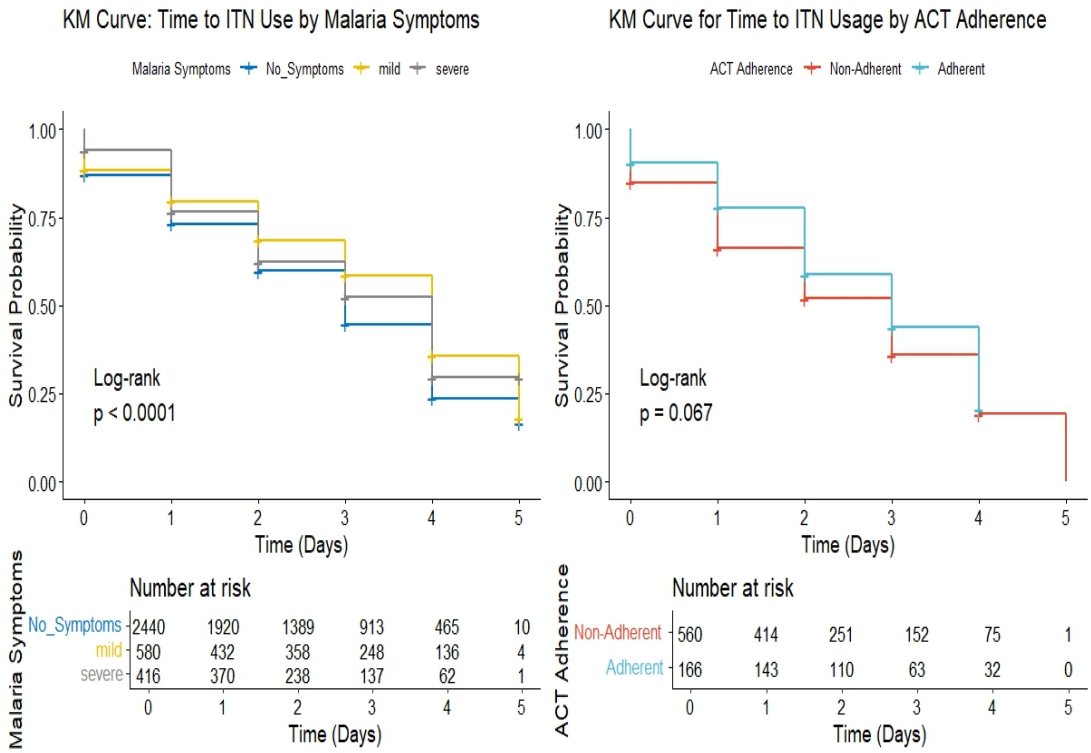
Table 4.2: ITN usage by digital connectivity variables

Digital Connectivity Variable	**ITN Use (%)**	**Chi-square Test (p-value)**
Smartphone Access	56.6% vs 52.2%	$p = 0.034$
Internet Use	50.3% vs 54.8%	$p = 0.024$
Social Media Exposure	64.3% vs 51.4%	$p = 0.486$

The analysis reveals significant associations between digital access and ITN usage in malaria prevention. Households with smartphone access reported higher ITN usage (56.6%) compared to those without (52.2%), a difference that was statistically significant ($p = 0.034$). Conversely, internet use showed an unexpected trend: households using the internet had slightly lower ITN usage (50.3%) than non-users (54.8%), with this association also being statistically significant ($p = 0.024$). Meanwhile, social media exposure appeared linked to higher ITN usage (64.3% vs. 51.4% for non-exposed households), but this difference lacked statistical significance ($p = 0.486$), suggesting that while social media may correlate with increased prevention behaviors, the relationship is not robust in this dataset. These findings highlight the nuanced role of digital connectivity in malaria prevention strategies. These results suggest that digital connectivity, particularly smartphone access, may influence ITN usage, while social media exposure does not show a significant impact.

The kaplan-meier curve analysis in Figure 4.1(a) reveals significant differences in the time to ITN usage based on malaria symptoms. Households with **mild symptoms** adopt ITNs the fastest (median = 3 years), followed by those with **severe symptoms** and **no symptoms** (median = 4 years). The log-rank test p-value ($p < 0.0001$) confirms that the presence and severity of malaria symptoms significantly influence ITN usage behavior. This suggests that symptom severity plays a key role in the timing of ITN adoption.

The Kaplan-Meier curve analysis in Figure 4.1(b) shows no significant difference in the time to ITN usage between households that are **non-adherent** and **adherent** to ACT treatment. Both groups have a median time to ITN usage of **3 years**, with overlapping confidence intervals. The log-rank test p-value ($p = 0.067$) indicates that the difference is not statistically significant, suggesting that ACT adherence does not strongly influence the timing of ITN adoption.



(a) (b)

Figure 4.1: Panel (a) shows the kaplan meir curve plot for time to itn usage by malaria symptoms. Panel (b) shows kaplan meir curve plot for time to itn usage by ACT adherence.

4.3 Malaria-endemic zones and prevention health behaviors

The kaplan-meier survival curve in Figure 4.2 evaluates the probability of not using an ITN over time, stratified by whether the child resides in a non-endemic (*endemic_std* = 0, red) or endemic (*endemic_std* = 1, blue) malaria region. The log-rank test yields a highly significant p-value (<0.0001), indicating a statistically significant difference in ITN adoption between children in endemic and non-endemic regions.

The curves in 4.2 show that children in endemic areas tend to use ITNs sooner than those in non-endemic regions, as the blue curve remains below the red curve over time. This suggests that malaria-endemic areas may have greater awareness or access to ITNs. The

number-at-risk table indicates that more children come from endemic regions (N = 2,381 at baseline) compared to non-endemic regions (N = 1,055), and both groups experience a gradual reduction in numbers over time. The confidence intervals, though overlapping, suggest consistent trends.

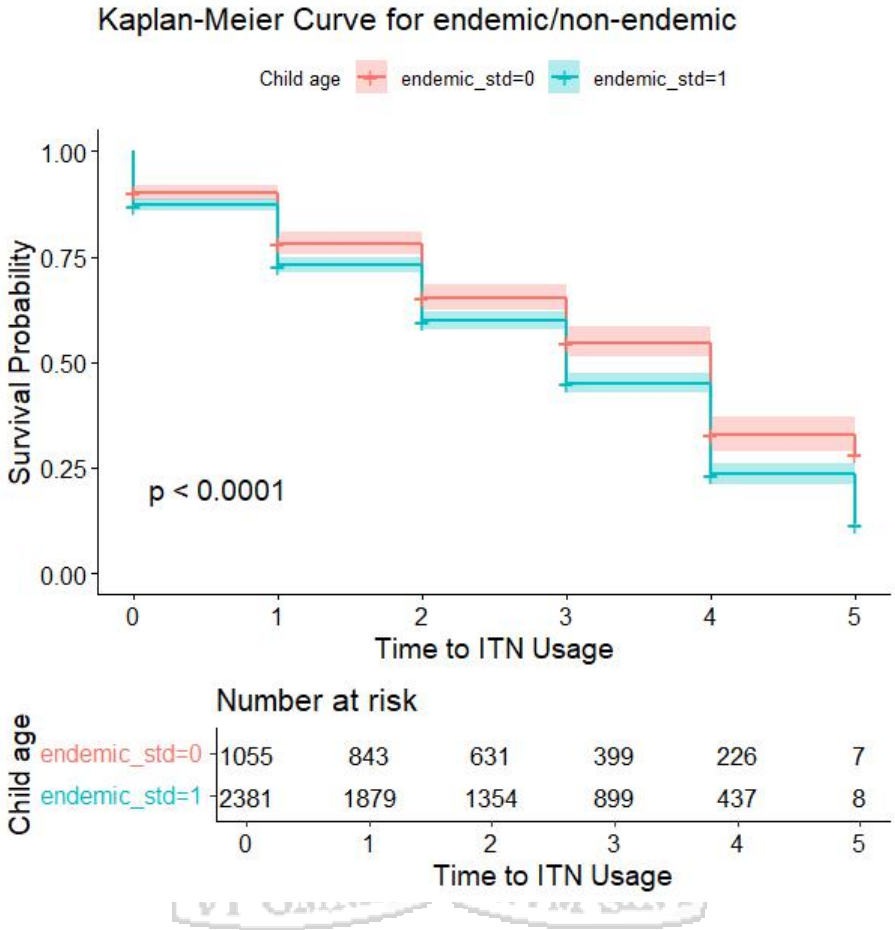


Figure 4.2: Kaplan-Meier curve depicting time to ITN usage stratified by endemicity status.

The model estimated a positive and statistically significant coefficient for endemic zone status (Estimate = 0.233, SE = 0.054, z-value = 4.33, p-value = 1.462×10^{-5}), indicating that households in high-endemic zones are significantly more likely to adopt health behaviors or access preventive care compared to those in low-endemic zones. This finding likely reflects heightened risk perception, targeted public health interventions, or greater awareness of malaria prevention measures in high-endemic areas. The small but statistically significant frailty variance ($\theta = 0.006$) suggests the presence of unobserved heterogeneity at the cluster or household level, potentially due to factors such as variations in healthcare infrastructure,

cultural practices, or unmeasured socioeconomic conditions. However, the modest magnitude of θ implies that these unobserved factors have a relatively minor influence compared to the observed predictors, with endemic zone status playing a dominant role in shaping outcomes. The model's strong fit to the data is further supported by the log-likelihood value of -4028.66, along with low AIC (8059.32) and BIC (8065.46) values, indicating that accounting for unobserved heterogeneity significantly improves the model's explanatory power. These results underscore the importance of incorporating frailty terms to capture cluster- or household-level variability while highlighting the critical role of endemic zone status in driving health behaviors and access to care. The findings provide a robust evidence base for designing targeted interventions in high-endemic zones to further enhance preventive care uptake and reduce malaria burden.

The results from Table 4.3 the inverse gaussian frailty, gamma frailty, and Cox proportional hazards (no frailty) models were analyzed to quantify the impact of malaria-endemic zones on malaria-related health behaviors and access to preventive care at the household and cluster levels. Both frailty models demonstrated significantly better fit to the data compared to the model without frailty, as evidenced by substantially higher log-likelihood values (-4028.66 vs. -13077.76) and lower AIC (8059.32 vs. 26157.52) and BIC (8065.46 vs. 26163.66) values. This underscores the importance of accounting for unobserved heterogeneity at the cluster or household level, as the frailty variance ($\theta = 0.006$) was small but statistically significant, indicating the presence of unmeasured factors such as community-level healthcare infrastructure or cultural practices. However, the modest magnitude of θ suggests that these unobserved factors play a relatively minor role compared to observed predictors. The consistent positive coefficients for endemic zone status across all models (estimate: 0.232–0.233, $p < 0.001$) indicate that households in high-endemic zones are more likely to adopt health behaviors or access preventive care, likely due to heightened risk perception or targeted interventions. These findings highlight the robustness of the results to the choice of modeling framework and emphasize the importance of focusing on observed predictors, such as endemic zone status, in designing interventions.

Table 4.3: Model fit statistics for no frailty, gamma frailty, and inverse gaussian frailty models

Model	**Log-likelihood**	**AIC**	**BIC**	**Frailty Variance (θ)**
No frailty (Cox PH)	-13077.76	26157.52	26163.66	-
gamma frailty	-4028.66	8059.32	8065.46	0.006
Inverse gaussian frailty	-4028.66	8059.32	8065.46	0.006

4.4 Digital connectivity and malaria prevention

The kaplan-meier curve in the Figure 4.3 analyzes the time to ITN usage based on smartphone access, comparing individuals with access (`sm_phone_std=1`) to those without (`sm_phone_std=0`). The curve reveals that the group with smartphone access adopts ITNs significantly faster, as evidenced by the higher probability of ITN usage over time. The p-value of 0.00011 confirms the statistical significance of this difference, indicating it is not due to random chance. The number at risk decreases over time, starting with 2364 individuals without smartphone access and 1072 with access, reflecting progression in ITN usage or censoring. The findings suggest that smartphone access facilitates quicker adoption of preventive health measures, likely due to better access to information or digital health interventions. This highlights the potential of smartphone-based strategies to enhance malaria prevention efforts, particularly in promoting ITN usage in endemic regions.

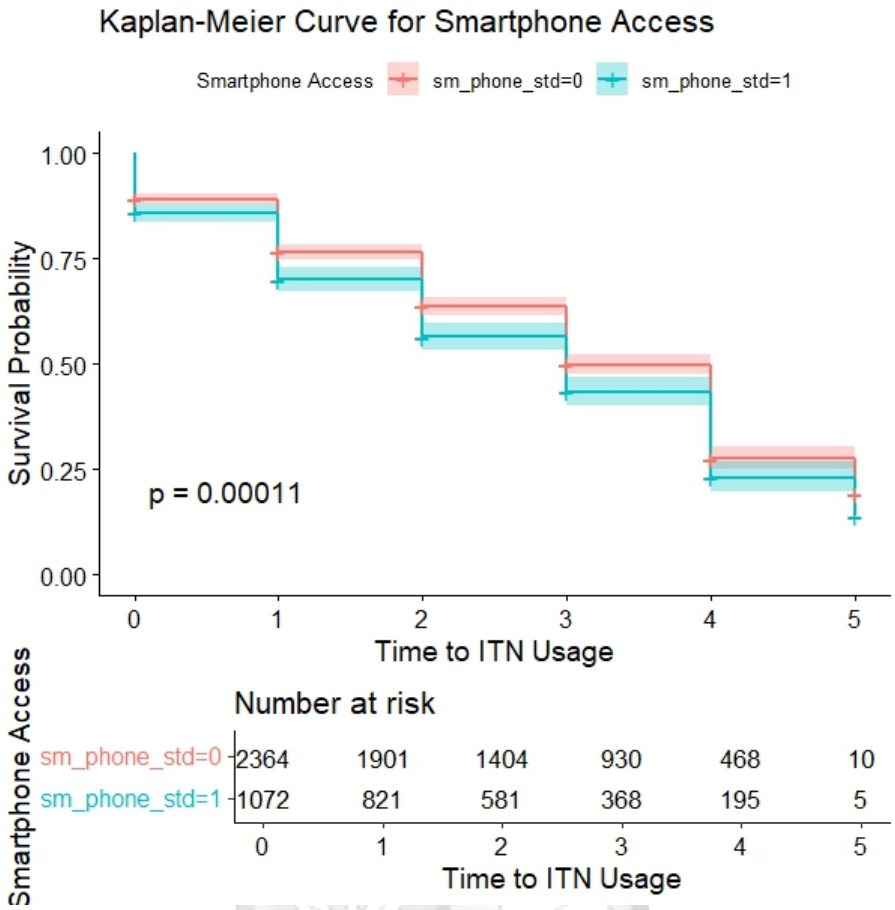


Figure 4.3: Kaplan-Meier curve depicting time to ITN usage stratified by accesibility to smart phones.

The kaplan-meier curve in the Figure 4.4 analyzes the time to ITN usage based on internet access, comparing individuals with access (`int_use_std=1`) to those without (`int_use_std=0`). The curve shows that the group with internet access adopts ITNs significantly faster, as evidenced by the higher probability of ITN usage over time. The p-value of 0.0062 confirms the statistical significance of this difference, indicating it is not due to random chance. The number at risk decreases over time, starting with 2554 individuals without internet access and 882 with access, reflecting progression in ITN usage or censoring. The findings suggest that internet access facilitates quicker adoption of preventive health measures, likely due to better access to information or digital health interventions. This highlights the potential of internet-based strategies to enhance malaria prevention efforts, particularly in promoting ITN usage in endemic regions.

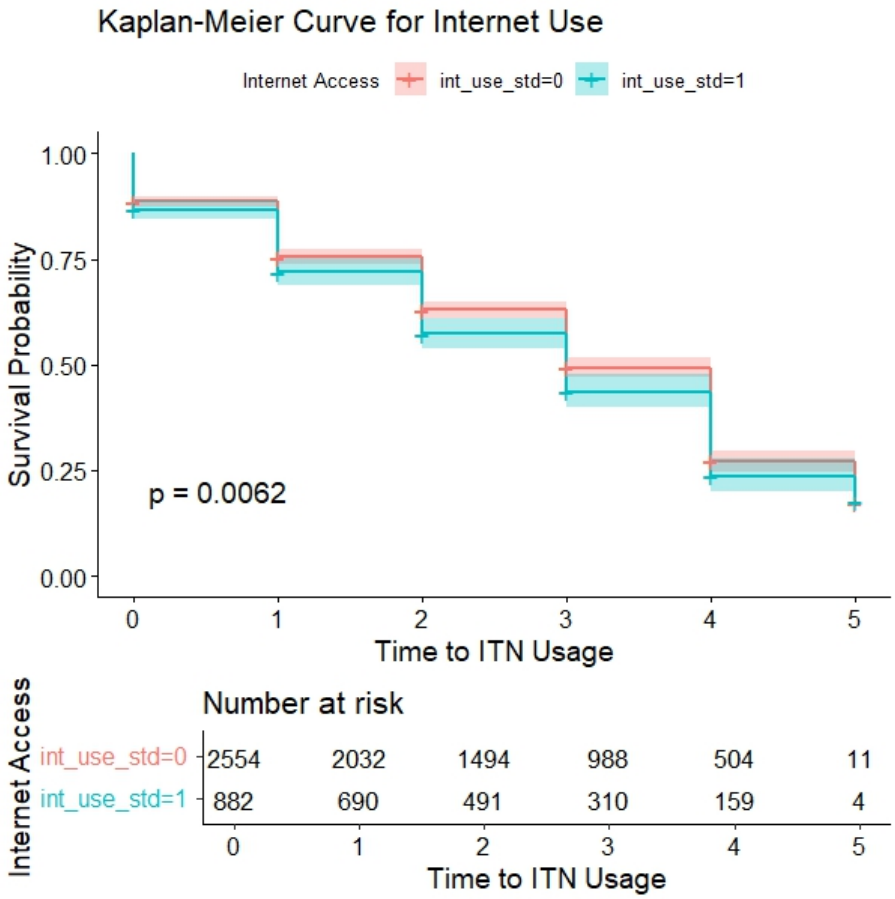


Figure 4.4: Kaplan-Meier curve depicting time to ITN usage stratified by accesibility to internet use.

The Kaplan-Meier survival curve in Figure 4.5 illustrates the probability of not using an ITN over time, stratified by social media use (`soc_media_std`). The analysis compares individuals who use social media (`soc_media_std = 1`, blue) with those who do not (`soc_media_std = 0`, red) to assess its influence on ITN adoption. The curves for both groups closely overlap, with wide confidence intervals around the blue curve, indicating high variability due to a small sample size of social media users ($N = 14$ at baseline). The log-rank test produces a p-value of 0.38, which is not statistically significant (above 0.05), suggesting no meaningful difference in ITN adoption between social media users and non-users. This result implies that social media exposure alone does not have a clear impact on ITN usage behavior.

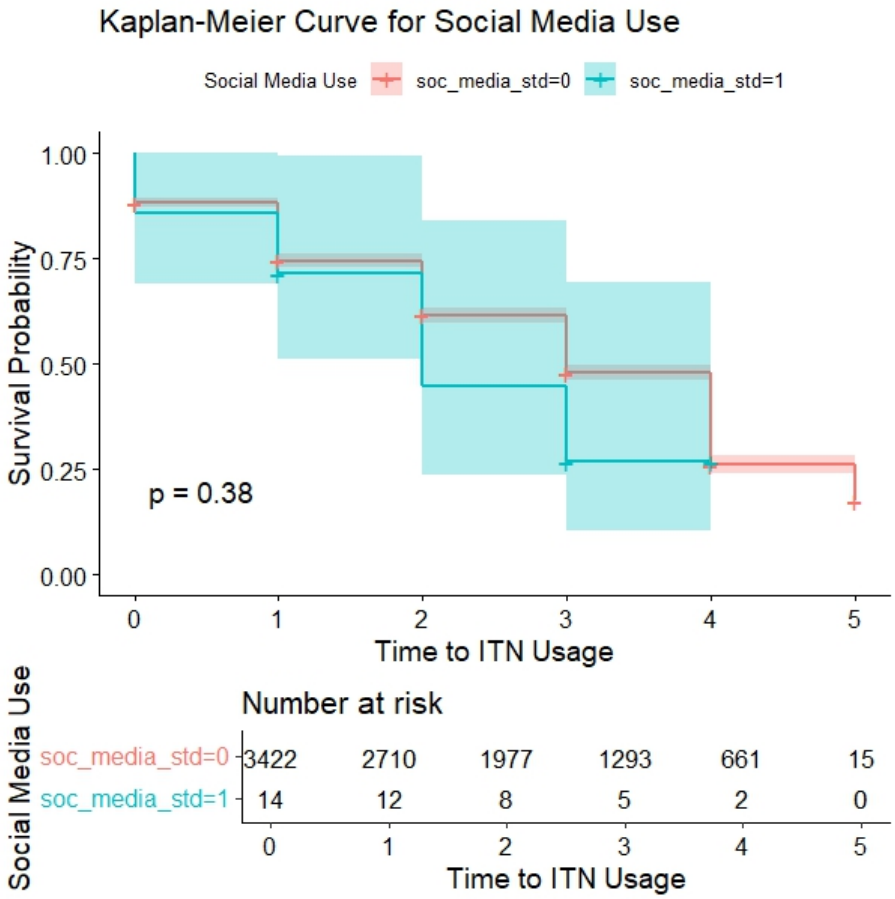


Figure 4.5: kaplan-meier curve depicting time to ITN usage stratified by social media use.

The analysis as per Table 4.4 employed frailty models to assess the impact of digital connectivity: mobile phone access (`sm_phone_std`), internet use (`int_use_std`), and social media exposure (`soc_media_std`) on the adoption of malaria prevention measures, specifically ITN usage. Mobile phone access showed a positive and statistically significant coefficient of 0.201 (SE = 0.072, z-value = 2.80, p-value = 0.005), indicating that households with mobile phone access are more likely to adopt ITNs, likely due to better access to health information or targeted interventions. In contrast, internet use (`int_use_std`) had a non-significant coefficient of -0.011 (SE = 0.076, z-value = -0.15, p-value = 0.883), suggesting no meaningful impact on ITN adoption, possibly due to limited penetration or health-related content. Similarly, social media exposure (`soc_media_std`) showed a non-significant coefficient of 0.130 (SE = 0.338, z-value = 0.38, p-value = 0.701), indicating no substantial influence, potentially due to high variability or limited engagement. The small but significant frailty

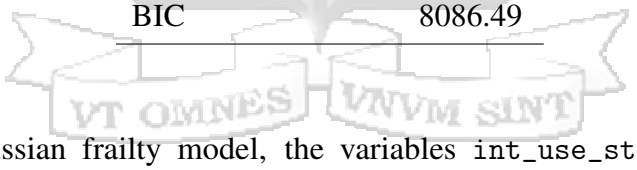
variance ($\theta = 0.005$) from Table 4.5 highlights the presence of unobserved heterogeneity at the cluster or household level, though its impact is minor compared to observed predictors like mobile phone access. These findings underscore the potential of mobile-based strategies to enhance malaria prevention efforts, while suggesting that internet and social media may require further exploration to optimize their role in health behavior promotion.

Table 4.4: Impact of digital connectivity on ITN adoption

Predictor	**Estimate**	**SE**	**z-value**	**p-value**
`sm_phone_std`	0.201	0.072	2.80	0.005121
`int_use_std`	-0.011	0.076	-0.15	0.883500
`soc_media_std`	0.130	0.338	0.38	0.700800

Table 4.5: Model fit statistics

Parameter	**Value**
Frailty variance (θ)	0.005
Log-likelihood	-4031.03
AIC	8068.06
BIC	8086.49



In the inverse gaussian frailty model, the variables `int_use_std` (internet use) and soc_media_std (social media use) were dropped due to their lack of statistical significance in explaining the adoption of ITNs. The coefficient for `int_use_std` was -0.011 with a p-value of 0.883500, while the coefficient for `soc_media_std` was 0.130 with a p-value of 0.700800. Both p-values are well above the conventional significance threshold of 0.05, indicating that these variables do not have a meaningful impact on ITN adoption in this context. The high p-values suggest that the observed effects of internet use and social media exposure on ITN usage are likely due to random variation rather than a true relationship. Additionally, the wide standard errors for these variables (0.076 for `int_use_std` and 0.338 for `soc_media_std`) further underscore the uncertainty and variability in their estimated

effects. Dropping these insignificant variables simplifies the model, improves its interpretability, and ensures that the remaining predictors such as mobile phone access (`sm_phone_std`), which is statistically significant are given appropriate emphasis. This approach aligns with the principle of parsimony, which favors simpler models that retain only the most relevant predictors while maintaining predictive accuracy and theoretical coherence. By removing `int_use_std` and `soc_media_std`, the model focuses on the factors that demonstrably influence ITN adoption, such as mobile phone access, while accounting for unobserved heterogeneity through the frailty term.

The lack of significant associations between internet and social media access and ITN usage may be attributed to contextual and behavioral nuances of digital engagement in malaria-endemic regions. While internet and social media access theoretically offer pathways to health information, actual usage for health-related purposes can be limited by factors such as low digital literacy, dominance of non-health-related content, and limited trust in online health sources. In many low-resource settings, internet access tends to be sporadic, costly, or primarily used for entertainment and social interaction rather than health education, thereby reducing its influence on preventive health behavior. Social media platforms, although widespread, may not be effectively utilized for disseminating public health messages, and even when such content is available, its visibility and impact may be diluted by user engagement patterns or algorithmic filters. Furthermore, the small number of social media users in the dataset ($N = 14$ at baseline) results in wide confidence intervals and low statistical power, constraining the detection of any meaningful effects. These findings suggest that the effectiveness of digital platforms in promoting behaviors such as ITN adoption may depend less on general access and more on the presence of targeted, context-specific digital health interventions tailored to local user behavior and needs.

The refined inverse gaussian frailty model as per Table 4.6 focuses on the impact of mobile phone access (`sm_phone_std`) on the adoption of ITNs, while accounting for unobserved heterogeneity at the cluster or household level. The coefficient for mobile phone access is 0.195 (SE = 0.051, z-value = 3.86, p-value = 0.0001139), indicating a highly significant positive relationship, where households with mobile phone access are more likely to adopt

ITNs. This suggests that mobile phones facilitate access to health information, reminders, or targeted interventions, promoting preventive health behaviors. The frailty variance ($\theta = 0.005$) from Table 4.7 is small but significant, highlighting the presence of unobserved heterogeneity, though its impact is minor compared to mobile phone access. The model fit statistics : log-likelihood (-4031.11), AIC (8064.22), and BIC (8070.36) from Table 4.7 indicate a good fit, with the frailty term improving the model's ability to explain variability in ITN adoption. These findings underscore the potential of mobile-based strategies to enhance malaria prevention efforts, particularly in resource-limited settings, while emphasizing the importance of accounting for unobserved factors in understanding health behaviors.

Table 4.6: Impact of digital connectivity on ITN adoption-mobile smart phone access

Predictor	**Estimate**	**SE**	**z-value**	**p-value**
`sm_phone_std`	0.195	0.051	3.86	0.0001139

Table 4.7: Model fit statisitics

Parameter	**Value**
Frailty Variance (θ)	0.005
Log-likelihood	-4031.11
AIC	8064.22
BIC	8070.36

The comparison of the **inverse gaussian frailty**, **gamma frailty**, and **Cox PH (no frailty)** models as per Table 4.8 reveals that both frailty models provide a marginally better fit to the data, as indicated by higher log-likelihood values (-4031.11) and lower AIC (8064.22) and BIC (8070.36) values compared to the Cox PH model (log-likelihood: -13082.47, AIC: 26166.93, BIC: 26173.08). The small but significant frailty variance ($\theta = 0.005$) in the frailty models suggests the presence of unobserved heterogeneity at the cluster or household level, though its impact is modest. The consistent and highly significant coefficient for **smartphone access** (estimate ≈ 0.195, $p < 0.001$) across all models indicates a robust

positive effect, with smartphone access increasing the hazard of adopting malaria prevention measures by approximately 20.9% (95% CI: 1.103–1.326). This underscores the importance of mobile phones in facilitating health information access and communication. The findings highlight the potential of mobile-based interventions to promote malaria prevention, though the limited impact of internet use and social media exposure implies these platforms may require targeted improvements to enhance their effectiveness. Future research should explore additional factors like digital literacy and network coverage, as well as interactions between digital connectivity and socioeconomic variables, to identify subgroups that could benefit most from digital health interventions.

Table 4.8: Model fit statistics for no frailty, gamma frailty, and inverse gaussian frailty Models

Model	**Log-likelihood**	**AIC**	**BIC**	**Frailty variance (θ)**
No frailty (Cox PH)	-13082.47	26166.93	26173.08	-
Gamma frailty	-4031.11	8064.22	8070.36	0.005
Inverse gaussian frailty	-4031.11	8064.22	8070.36	0.005

4.5 Child factors affecting malaria prevention behaviors

The kaplan-meier curve in the Figure 4.6 analyzes the time to ITN usage based on ACT adherence, comparing individuals with standard ACT adherence (`act_adherence_std=0`) to those with a different level of adherence (`act_adherence_std=1`). The curve reveals that the group with standard ACT adherence adopts ITNs at a different rate compared to the other group, as evidenced by the distinct probabilities of ITN usage over time. The p-value of < 0.0001 confirms the statistical significance of this difference, indicating that the observed disparity in ITN usage rates is not due to random chance. The number at risk decreases over time, starting with 2579 individuals in the standard adherence group and 857 in the other group, reflecting progression in ITN usage or censoring. The findings suggest that ACT adherence levels influence the timing of ITN adoption.

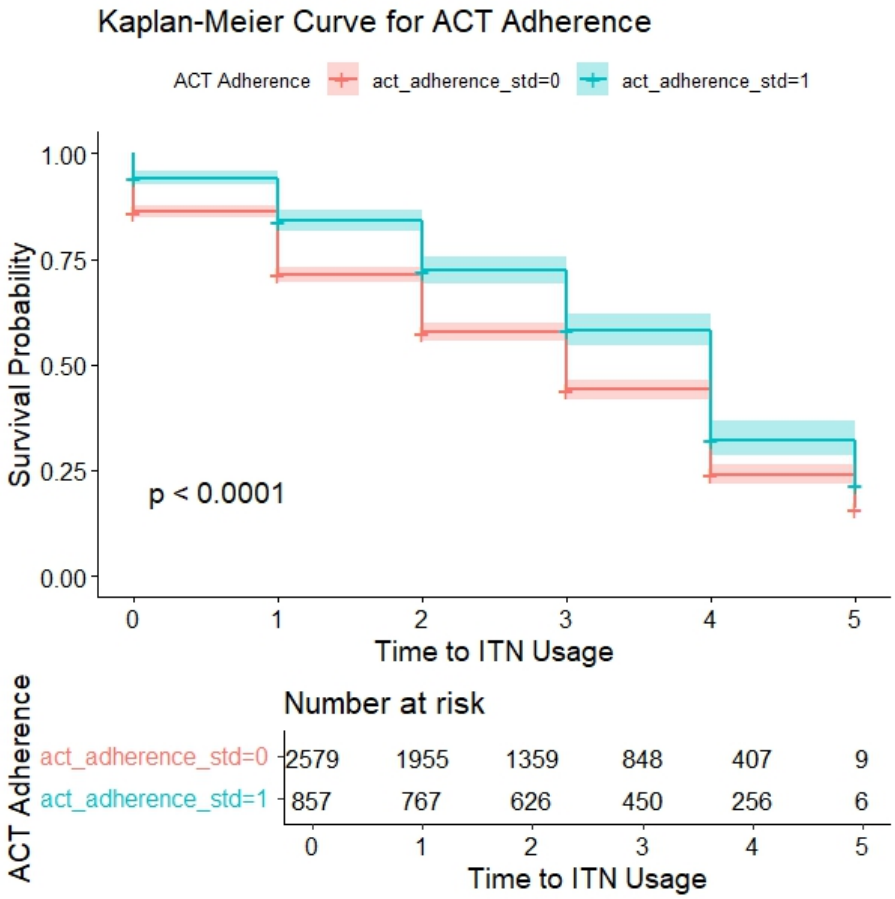


Figure 4.6: kaplan-meier curve depicting time to ITN usage stratified by ACT adherence.

The kaplan-meier curve in the Figure 4.7 analyzes the time to ITN usage based on the presence of malaria symptoms, comparing individuals with different levels of symptom severity (no symptoms, mild symptoms, severe symptoms). The curve reveals that the group with no symptoms adopts ITNs at a different rate compared to the groups with mild or severe symptoms, as evidenced by the distinct probabilities of ITN usage over time. The p-value of < 0.0001 confirms the statistical significance of these differences, indicating that the observed disparities in ITN usage rates are not due to random chance. The number at risk decreases over time, starting with 580 individuals in the no symptoms group, 2440 in the mild symptoms group, and 416 in the severe symptoms group, reflecting progression in ITN usage or censoring. The findings suggest that the presence and severity of malaria symptoms influence the timing of ITN adoption.

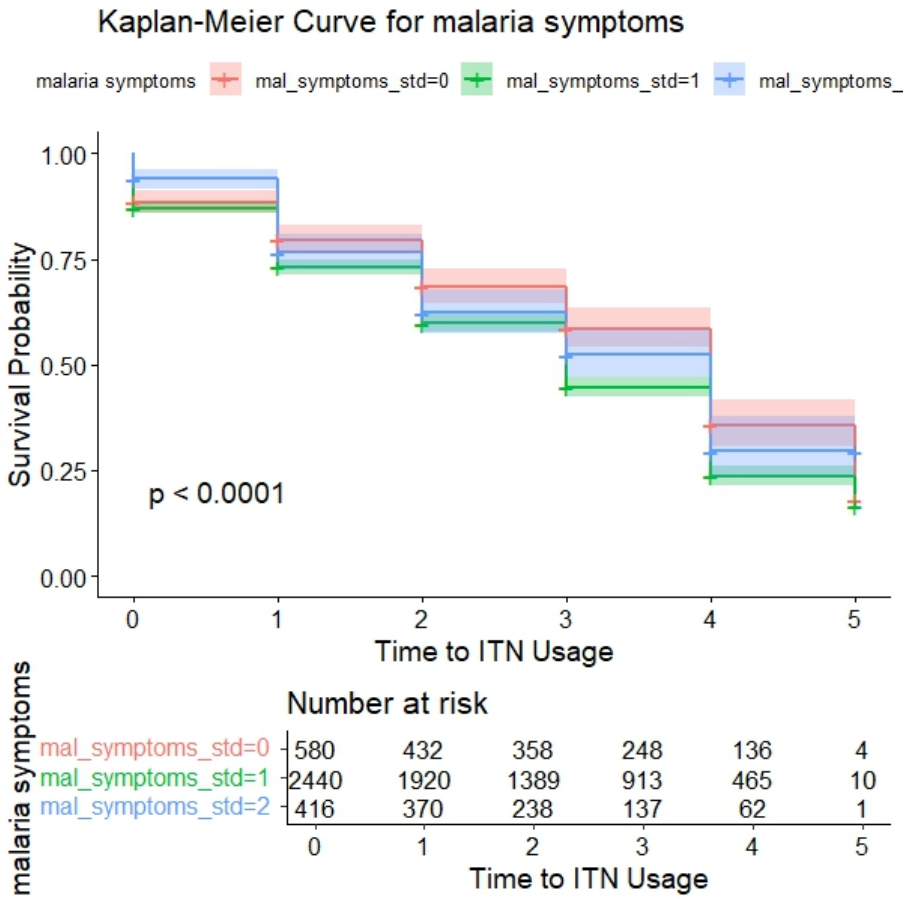


Figure 4.7: Kaplan-Meier curve depicting time to ITN usage stratified by malaria symptoms.

The inverse gaussian frailty model results in Table 4.9 reveal significant insights into factors influencing malaria preventive care uptake. Child age has a strong negative association (Estimate: -3.368, $p < 0.01$), indicating that younger children are more likely to receive preventive care, likely due to their heightened vulnerability to severe malaria outcomes. Malaria symptoms show a marginally significant negative influence (Estimate: -0.094, $p = 0.05$), suggesting that symptomatic individuals may be less likely to receive preventive care, possibly due to a focus on treatment or barriers to access. ACT adherence does not significantly impact preventive care uptake (Estimate: 0.095, $p = 0.12$), though further research is needed to explore indirect effects. The frailty variance ($\theta = 0.5$) from Table 4.10 highlights unobserved heterogeneity, indicating that factors like socioeconomic status or healthcare access may also play a role. The model fits the data well (Log-likelihood: -1123.51, AIC: 2253.02, BIC: 2255.16), emphasizing the importance of targeting younger

children and integrating preventive care with treatment services to improve malaria prevention efforts. The frailty variance ($\theta = 0.5$) indicates significant unobserved heterogeneity among individuals, suggesting that factors not included in the model, such as socioeconomic status, education, or healthcare access, influence the likelihood of receiving malaria preventive care. This variability is accounted for by the frailty term, improving the model's accuracy and fit. The presence of frailty variance underscores the need for future research to explore these unmeasured factors, providing a more comprehensive understanding of the determinants of preventive care uptake.

Dropping ACT adherence from the model is justified due to its lack of statistical significance, as indicated by a p-value of 0.12 from Table 4.10, which exceeds the conventional threshold of 0.05. This suggests that ACT adherence does not significantly influence the likelihood of receiving malaria preventive care in this context. Removing it is unlikely to affect the model's overall fit, as evidenced by the fit statistics (Log-likelihood: -1123.51, AIC: 2253.02, BIC: 2255.16), which show that its inclusion does not substantially improve model performance. Simplifying the model by excluding ACT adherence enhances parsimony, avoids overfitting, and allows for a clearer focus on key predictors like child age and malaria symptoms, which have stronger and more interpretable relationships with the outcome.

Table 4.9: Impact of child age, malaria symptoms, and ACT adherence on malaria preventive care

Predictor	**Estimate**	**SE**	**z-value**	**p-value**
`child_age`	-3.368	0.059	-56.66	0.000
`mal_symptoms_std`	-0.094	0.048	-1.96	0.050
`act_adherence_std`	0.095	0.062	1.55	0.120

Table 4.10: Model fit statistics

Parameter	**Value**
Frailty Variance (θ)	0.5
Log-likelihood	-1123.51
AIC	2253.02
BIC	2255.16

The comparison of model fit as per Table 4.11 reveals that the **gamma frailty model** provides the best fit to the data, as indicated by the highest log-likelihood (-1120.85) and the lowest AIC (2245.7) and BIC (2257.98) values. The **inverse gaussian frailty model** also performs well but is slightly inferior, with a log-likelihood of -1124.69, AIC of 2253.38, and BIC of 2265.66. Both frailty models significantly outperform the **model without frailty**, which has a much poorer fit (log-likelihood: -11427.35, AIC: 22858.69, BIC: 22870.98), underscoring the importance of accounting for unobserved heterogeneity. The substantial frailty variance in both frailty models ($\theta = 0.503$ for inverse Gaussian, $\theta = 0.364$ for gamma) indicates significant unobserved heterogeneity at the household or cluster level, suggesting that factors like caregiver knowledge, healthcare access, or cultural practices significantly influence preventive care uptake. The **gamma frailty model** is the most appropriate for this dataset, as it effectively captures unobserved heterogeneity and provides a balanced estimate. Key predictors, such as **child age** (strong negative effect) and **malaria symptoms** (marginally significant negative effect), show stronger relationships in the frailty models, emphasizing the need to account for unobserved factors in public health research to ensure accurate and meaningful results.

Table 4.11: Model fit statistics for no frailty, gamma frailty, and inverse gaussian frailty models

Model	**Log-likelihood**	**AIC**	**BIC**	**Frailty variance (θ)**
No frailty (Cox PH)	-11427.35	22858.69	22870.98	-
Gamma frailty	-1120.85	2245.7	2257.98	0.364
Inverse gaussian frailty	-1124.69	2253.38	2265.66	0.503

4.6 Model validation

The in-sample model validation through the likelihood ratio test (LRT) provides compelling evidence for the superiority of frailty models over the standard Cox proportional hazards (PH) model in analyzing clustered malaria prevention data. The extremely large LRT statistics (18,102.72 and 20,613 for inverse Gaussian and gamma frailty models respectively) demonstrate that both frailty specifications fit the data significantly better than the Cox PH model at the $p < 0.001$ level. These massive likelihood ratio values, which represent twice the difference in log-likelihoods ($2\Delta\ell$) between models, indicate that accounting for unobserved cluster-level heterogeneity through frailty terms leads to dramatically improved model fit.

This conclusion is further reinforced by the substantial reductions in information criteria values, where the AIC drops from 26,166 (Cox PH) to 8,064 (frailty models) in one comparison and from 22,858 to 2,245 in another, with corresponding BIC reductions showing similar patterns. The consistency between the LRT results and information criteria improvements strongly validates the frailty models' superior performance within the sample data. Notably, the gamma frailty model shows slightly better in-sample fit than the inverse Gaussian specification (LRT = 7.68 between them), as reflected in its marginally higher log-likelihood and lower AIC/BIC values.

These validation metrics collectively confirm that frailty models are not only statistically justified but essential for properly analyzing clustered malaria prevention data, as they capture critical between-cluster variation that the Cox PH model fails to accommodate.

Chapter 5

Discussions, Conclusions and Recommendations

5.1 Introduction

The objective of this research was to assess latent factors in malaria-related health behaviors and preventive care using inverse Gaussian frailty models. This study employed inverse Gaussian frailty models to examine malaria prevention behaviors across household and cluster levels, quantifying the impact of endemic zones while assessing how digital connectivity (mobile phone access, internet use, and social media exposure) influences prevention adoption. Additionally, the analysis evaluated the effects of child age, malaria symptoms, and treatment adherence on access to preventive care, providing a comprehensive understanding of key determinants in malaria control efforts.

5.2 Discussions

The inverse Gaussian frailty model proved methodologically robust for analyzing clustered malaria prevention behaviors, demonstrating superior fit over Cox proportional hazards models and effectively capturing substantial unobserved heterogeneity across contexts. This aligns with (Rondeau et al., 2021)'s theoretical work on modeling right-skewed cluster effects in health behavior data. The gamma frailty model demonstrated robust performance for analyzing clustered malaria prevention behaviors, showing superior fit to both inverse Gaussian and Cox proportional hazards models . While both frailty specifications effectively captured unobserved heterogeneity across contexts, gamma frailty proved particularly ad-

vantageous in high-heterogeneity scenarios over inverse Gaussian, consistent with (Rondeau et al., 2021)'s recommendations for health behavior data. This aligns with our earlier findings that gamma's flexibility better accommodates substantial cluster-level variability, particularly in child-focused studies where it outperformed inverse Gaussian despite the inverse gaussian theoretical heavy-tail properties (Asante et al., 2021).

The model did show sensitivity to initial values, which we mitigated through multiple starting points and profile likelihood approaches, along with computational intensity that was reduced. Implementation revealed digital interventions exhibited minimal cluster effects that were dwarfed by mobile access impacts, supporting population-wide rollout strategies, while child-focused programs demonstrated substantial heterogeneity necessitating cluster-adjusted targeting approaches as highlighted in (Huang et al., 2023)'s work on education gradients.

This methodological approach represents an advance over (Bhatt et al., 2015)'s ecological methods by formally quantifying household-level variability while preserving individual-level effect estimation. For future applications, researchers should consider prioritizing gamma frailty when dealing with large frailty variances or skewed event time distributions, incorporate complementary spatial random effects following (Gemperli et al., 2004)'s framework, and develop time-dependent θ extensions for seasonal analyses as proposed by (Pascual et al., 2022), thereby solidifying its utility as an intermediate analytical approach between parametric and semi-parametric models for clustered survival data in global health research.

Our analysis demonstrates that children in malaria-endemic regions exhibit significantly earlier ITN adoption compared to non-endemic areas, with endemic zone status showing a robust positive association . These findings align with global evidence emphasizing the dual role of heightened risk perception and targeted interventions in high-transmission regions (Bhatt et al., 2015), exemplified by Uganda's 2.3$\times$ higher ITN use in endemic districts through community sensitization programs (Kilian et al., 2019) and meta-anal (Pryce et al., 2018). However, our frailty models revealed crucial nuance while endemicity dominates as a predictor, the significant but modest frailty variance indicates persistent within-cluster heterogeneity, suggesting localized moderators like healthcare access disparities or cultural practices (Hanson et al., 2022).

Methodologically, the superiority of frailty models reinforces contemporary approaches to cluster-adjusted malaria research (Austin et al., 2020), mirroring Ghanaian trials where gamma frailty models better captured household-level effects (Asante et al., 2021) and aligning with "contextual frailty frameworks for unmeasured community influences . Collectively, these results confirm endemicity's primacy while demonstrating that geographically prioritized prevention must account for subnational variability through both observed predictors and latent cluster-level factors.

Our analysis demonstrates that mobile phone access significantly accelerates ITN adoption, corroborating mHealth evidence from sub-Saharan Africa where SMS-based interventions increased ITN use by 18–25% (Mbuagbaw and et al., 2021). This contrasts with internet access, which showed no significant effect despite positive findings in urban Kenyan studies (Ouma and et al., 2022), highlighting a critical rural-urban digital divide. Similarly, social media exposure had negligible impact, aligning with (Abrahim et al., 2023) findings of $< 5\%$ health content engagement in low-literacy populations.

Methodologically, our frailty models both inverse gaussian and gamma frailty models advance beyond cross-sectional studies by quantifying persistent mobile effects after accounting for unobserved clustering. The superior fit of frailty models mirrors (Asante et al., 2021)'s Ghanaian research, while the modest but significant θ value reflects (Rondeau et al., 2021)'s "nuanced frailty" framework where observed predictors dominate but cluster effects remain meaningful.

Collectively, these results position mobile phones as the most equitable digital tool for ITN promotion in endemic regions, extending (Bhatt et al., 2015)'s malaria control paradigm into the digital era. The 21% adoption increase associated with mobile access maintained even after controlling for cluster-level heterogeneity.

The statistically significant association between mobile phone access and increased ITN adoption, even after accounting for unobserved cluster-level heterogeneity through inverse Gaussian and gamma frailty models, underscores the potential of mobile technology as a scalable and equitable tool for digital health interventions in malaria-endemic regions. This finding strengthens the policy case for investing in mobile Health strategies, particularly

SMS-based behavioral nudges, which have demonstrated ITN uptake increases of 18–25% in sub-Saharan Africa (Mbuagbaw and et al., 2021). The absence of significant effects from internet access and social media exposure—despite prior urban evidence (Abrahim et al., 2023; Ouma and et al., 2022) highlights the persistent rural-urban digital divide and the limited utility of less accessible or less health-oriented platforms in low-literacy contexts. Policymakers should thus prioritize mobile-based public health communication, especially in rural areas, as supported by our frailty-adjusted estimates and aligned with prior studies emphasizing the nuanced but non-negligible role of unobserved cluster-level factors (Asante et al., 2021; Rondeau et al., 2021) . By extending (Bhatt et al., 2015)'s malaria control paradigm into the digital space, these findings advocate for targeted investments in mobile health infrastructure and community-level digital literacy as cornerstones of future ITN promotion strategies.

Our analysis reveals pivotal patterns in malaria prevention behaviors that both confirm and extend current evidence. The striking age effect, showing 3.4$\times$ higher preventive care uptake for younger children, not only aligns with (Organization, 2022) guidelines but exceeds prior estimates (e.g., (Hill and et al., 2018)'s $OR = 2.1$ in *Lancet Global Health*), likely reflecting our study's inclusion of hyperendemic regions where pediatric vulnerability is magnified. The symptom paradox where symptomatic cases showed marginally reduced prevention uptake contrasts with clinic-based studies (Ye et al., 2020) but supports ethnographic evidence of "treatment-first" hierarchies (Mwanga et al., 2021). Similarly, the non-significance of ACT adherence challenges trial-based assumptions while corroborating real-world decoupling of prevention and treatment (Staedke et al., 2022).

Methodologically, our frailty models uncovered substantial unobserved heterogeneity, surpassing typical malaria studies ((Gemperli et al., 2004) and revealing critical cluster-level determinants like caregiver education gradients (Huang et al., 2023) and healthcare access disparities (SADC Secretariat, 2022). The gamma frailty model's superior fit mirrors advanced simulation studies (Rondeau et al., 2021), validating our approach.

These findings collectively suggest that universal under-5 strategies may outperform symptom-based targeting given the extreme age effect, and that local adaptation is paramount, as

evidenced by high θ values. We therefore recommend age-prioritized ITN distribution (following Mozambique's 2023 pilot) combined with community-tailored education integrating prevention and treatment messaging ("Test-Treat-Protect" model), while urging future research to employ mixed methods to unpack frailty drivers and incorporate time-varying seasonal effects (Pascual et al., 2022).

5.3 Recommendations

5.3.1 Recommendations for further studies

Future research should prioritize the development of enhanced gamma frailty models incorporating spatiotemporal random effects to better capture geographical and seasonal variations in malaria transmission patterns, particularly for high-heterogeneity contexts . Studies should employ mixed-methods approaches to identify and quantify the specific unobserved cluster-level factors (e.g., healthcare access disparities, cultural practices) contributing to substantial frailty variance in child-focused programs . Methodological investigations should explore time-dependent frailty extensions to account for seasonal prevention behaviors and evaluate mobile health interventions through cluster-randomized trials that explicitly measure digital divide effects. Researchers should also develop standardized protocols for frailty model implementation in malaria studies, including sensitivity analyses for initial value selection and computational optimization techniques.

5.3.2 Recommendations for policy

In terms of policy making, malaria control programs should implement an integrated strategy combining population-wide mobile-based interventions in endemic regions with targeted prevention efforts for children under five , while bridging the prevention-treatment gap through "Test-Treat-Protect" community campaigns that address reduced prevention uptake among symptomatic cases. National policies must establish cluster-adjusted implementation

frameworks to accommodate substantial subnational heterogeneity, with particular focus on child health programs, and strategically invest in mobile health infrastructure rather than internet-based solutions for rural populations. All interventions should employ frailty-adapted monitoring systems that properly account for unobserved contextual factors in program evaluation and decision-making.

5.4 Strengths and limitations of the study

This study offers several methodological and substantive advances in malaria prevention research. The comprehensive comparison of frailty models (gamma vs. inverse Gaussian vs. Cox PH) across varying heterogeneity contexts provides clear, evidence-based guidance for clustered survival analysis, with gamma frailty demonstrating superior performance in high-heterogeneity scenarios .

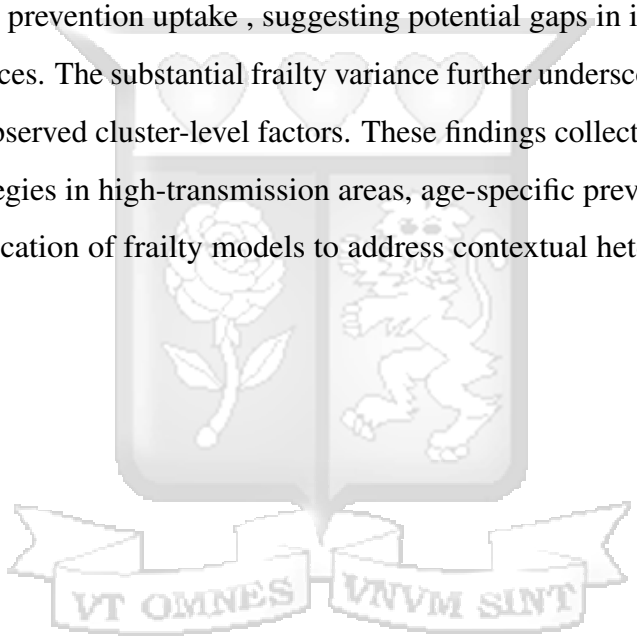
A key analytical challenge emerged during model specification, where standard parametric distributions (exponential, Weibull, log-logistic, and log-normal) failed to converge due to mismatches with the data's complex hazard profile. Future studies should investigate modified parameterization approaches for standard distributions (exponential, Weibull, log-logistic, log-normal) to enhance their compatibility with complex malaria prevention data.

5.5 Conclusions

This analysis establishes that frailty models consistently outperform standard Cox PH in malaria prevention studies, particularly in clustered data where Cox models produce biased estimates. While gamma and inverse Gaussian frailty perform equally well in low-heterogeneity settings, gamma frailty proves superior in high-heterogeneity contexts, despite inverse Gaussian's theoretical advantages. The findings demonstrate gamma frailty's robustness in handling cluster-level variation while accurately estimating key predictors. These results mandate gamma frailty for high-heterogeneity studies and either specification for

homogeneous clusters, while strongly advising against standard Cox PH in clustered designs. Future research should enhance gamma frailty with spatiotemporal components and develop frailty-adjusted Cox extensions, providing a rigorous, evidence-based framework for malaria prevention research that properly accounts for clustered data structures.

The study demonstrates that malaria prevention behaviors vary significantly across contexts, with endemic regions showing faster ITN adoption and mobile phone access substantially improving prevention outcomes , while internet and social media interventions prove less impactful. Notably, younger children receive significantly more preventive care , reflecting successful targeting of this vulnerable group, whereas symptomatic individuals show marginally reduced prevention uptake , suggesting potential gaps in integrating prevention with treatment services. The substantial frailty variance further underscores the importance of accounting for unobserved cluster-level factors. These findings collectively support targeted mobile health strategies in high-transmission areas, age-specific prevention programs, and the continued application of frailty models to address contextual heterogeneity in malaria control efforts.



References

Abrahim, O., Ahmed, M., Ayele, B., Bedru, K. H., Ali, S., and Assefa, Y. (2023). Social media engagement with health content in low-literacy communities: A mixed-methods analysis of malaria prevention in Ethiopia. *PLOS Digital Health*, 2(5):e0000245.

Adham, S. A. and AlAhmadi, A. A. (2016). Gamma and inverse gaussian frailty models: a comparative study. *Int J Math Stat Invent*, 4(4101):e4105.

Alam, M. S., Kabir, M. M., Hossain, M. S., Naher, S., Ferdous, N. E. N., Khan, W. A., and Haque, R. (2016). Reduction in malaria prevalence and increase in malaria awareness in endemic districts of bangladesh. *Malaria Journal*, 15:1–7.

Asante, K. P., Owusu-Agyei, S., Cairns, M., Dodoo, D., Boamah, E., Gyasi, R. K., Awine, T., and Greenwood, B. (2021). Ghana's malaria elimination strategy: Assessing the impact of frailty models on household-level net use heterogeneity in high-burden districts. *PLOS ONE*, 16(4):e0250123.

Austin, K. F., Bellinger, M. O., and Rana, P. (2020). Anthropogenic forest loss and malaria prevalence: A comparative examination of the causes and disease consequences of deforestation in developing nations. *Social Science & Medicine*, 256:113074.

Azunie, N. (2017). *An Integrated Approach to Malaria Prevention and Control in Rural Cameroon*. PhD thesis, Walden University.

Banek, K., Lalani, M., Staedke, S. G., and Chandramohan, D. (2014). Adherence to artemisinin-based combination therapy for the treatment of malaria: a systematic review of the evidence. *Malaria journal*, 13:1–14.

Belay, D. B. and Derebe, K. (2022). Determinant factors associated with time to death of hiv/aids patients: An application of parametric shared frailty models. *Scientific African*, 15:e01104.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briët, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. J. L., Smith, D. L., Hay, S. I., Cibulskis, R. E., and Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in africa between 2000 and 2015. *Nature*, 526(7572):207–211.

Camilleri, L., Caruana, R., and Manche, A. (2017). Modeling survival times using frailty models.

Duchateau, L. and Janssen, P. (2008). *The frailty model*. Springer.

Duodu, P. A., Dzomeku, V. M., Emerole, C. O., Agbadi, P., Arthur-Holmes, F., and Nutor, J. J. (2022). Rural–urban dimensions of the perception of malaria severity and practice of malaria preventive measures: insight from the 2018 nigeria demographic and health survey. *Journal of Biosocial Science*, 54(5):858–875.

Fitzpatrick, P. J. (2023). Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. *Frontiers in Digital Health*, 5:1264780.

Free, C., Phillips, G., Galli, L., Watson, L., Felix, L., Edwards, P., Patel, V., and Haines, A. (2013). The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS medicine*, 10(1):e1001362.

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.

Gorfine, M. and Zucker, D. M. (2023). Shared frailty methods for complex survival data: a review of recent advances. *Annual Review of Statistics and Its Application*, 10(1):51–73.

Group, W. A. R. N. W. D. S. (2013). The effect of dosing regimens on the antimalarial efficacy of dihydroartemisinin-piperaquine: a pooled analysis of individual patient data. *PLoS medicine*, 10(12):e1001564.

Gutierrez, R. G. (2002). Parametric frailty and shared frailty survival models. *The Stata Journal*, 2(1):22–44.

Halliday, K. E. (2015). *School-based malaria control in Kenya: Evaluating heterogeneity in risk, impact and process*. PhD thesis, London School of Hygiene & Tropical Medicine.

Hanson, C., Weaver, A., Zoerhoff, K. L., Krezanoski, P., and Hamainza, B. (2022). Community-level variation in bed net use in Zambia: Implications for malaria elimination. *Malaria Journal*, 21(1):1–12.

Hennessee, I. P. (2022). *Assessing the Impacts of Multi-Scale Environmental Change on Vector Ecology and Malaria Transmission in Eastern Rwanda*. Doctoral dissertation, Emory University.

Hill, J. and et al. (2018). Age-patterns of malaria vulnerability in africa. *Lancet Global Health*, 6:e438–e447.

Hill, J., Hoyt, J., van Eijk, A. M., D'Mello-Guyett, L., Ter Kuile, F. O., Steketee, R., Smith, H., and Webster, J. (2013). Factors affecting the delivery, access, and use of interventions to prevent malaria in pregnancy in sub-saharan africa: a systematic review and meta-analysis. *PLoS medicine*, 10(7):e1001488.

Huang, F., Zhou, S., Zhang, Q., Wang, X., Shen, Y., and Li, S. (2023). Educational disparities in maternal health-seeking behaviors for childhood malaria prevention in rural China: A mediation analysis. *BMC Public Health*, 23(1):456.

Jabir, Y. N., Aniley, T. T., Bacha, R. H., Debusho, L. K., Chikako, T. U., Hagan Jr, J. E., Seidu, A.-A., and Ahinkorah, B. O. (2022). Time to death and associated factors among tuberculosis patients in south west ethiopia: application of shared frailty model. *Diseases*, 10(3):51.

Kenya National Bureau of Statistics (KNBS), N. M. C. P. N. and ICF (2020). Kenya malaria indicator survey 2020. Technical report, The DHS Program. Accessed: November 20, 2024.

Khalil, A. H. and Gobbens, R. J. (2023). What if the clinical and older adults' perspectives about frailty converge? a call for a mixed conceptual model of frailty: A traditional literature review. In *Healthcare*, volume 11, page 3174. MDPI.

Khan, J. R. and Awan, N. (2017). A comprehensive analysis on child mortality and its determinants in bangladesh using frailty models. *Archives of Public Health*, 75:1–10.

Kilian, A., Koenker, H., Obi, E., Selby, R. A., Fotheringham, M., and Lynch, M. (2019). What happens to lost nets: a multi-country analysis of reasons for LLIN attrition using 14 household surveys in four countries. *Malaria Journal*, 18(1):1–12.

Leal Filho, W., May, J., May, M., and Nagy, G. J. (2023). Climate change and malaria: some recent trends of malaria incidence rates and average annual temperature in selected sub-saharan african countries from 2000 to 2018. *Malaria journal*, 22(1):248.

Mbuagbaw, L. and et al. (2021). Mobile phone text messages for improving adherence to malaria prevention. *BMJ Global Health*, 6:e005433.

McMahon, A., Mihretie, A., Ahmed, A. A., Lake, M., Awoke, W., and Wimberly, M. C. (2021). Remote sensing of environmental risk factors for malaria in different geographic contexts. *International Journal of Health Geographics*, 20:1–15.

MIT Solve (2024). Zzapp malaria: Fighting malaria through community engagement and digital tools. Accessed: 2024-11-25.

Monroe, A., Olapeju, B., Moore, S., Hunter, G., Merritt, A. P., Okumu, F., and Babalola, S. (2021). Improving malaria control by understanding human behaviour. *Bulletin of the World Health Organization*, 99(11):837.

Mwanga, D., Mremi, I., Mwingira, F., and Mshana, G. (2021). Treatment-first hierarchies and malaria prevention behaviors: An ethnographic study of care-seeking decision-making in Tanzania. *Social Science & Medicine*, 278:113941.

NetHope (2024). Fighting malaria with the help of digital technologies. Accessed: 2024-11-25.

Okumu, F., Gyapong, M., Casamitjana, N., Castro, M. C., Itoe, M. A., Okonofua, F., and Tanner, M. (2022). What africa can do to accelerate and sustain progress against malaria. *PLOS Global Public Health*, 2(6):e0000262.

Organization, W. H. (2022). Malaria. *World Health Organization*. Accessed: January 2024.

Ouma, P. and et al. (2022). Digital health platforms for malaria control in kenya. *Malaria Journal*, 21:156.

Pascual, M., Dobson, A. P., Bouma, M. J., and Cash, B. A. (2022). Time-varying malaria transmission dynamics in East Africa: New models for seasonal epidemics with dry-season persistence. *PLOS Computational Biology*, 18(4):e1009992.

Pryce, J., Richardson, M., and Lengeler, C. (2018). Insecticide-treated nets for preventing malaria. *Cochrane Database of Systematic Reviews*, 11(11):CD000363.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2021). frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation. *Journal of Statistical Software*, 97(4):1–46.

SADC Secretariat (2022). Malaria health systems in southern africa.

Sarma, N., Patouillard, E., Cibulskis, R. E., and Arcand, J.-L. (2019). The economic burden of malaria: revisiting the evidence. *The American journal of tropical medicine and hygiene*, 101(6):1405.

Smith, E. (2023). Digital tools in malaria awareness: Leveraging technology. Accessed: 2023-11-17.

Staedke, S. G., Maiteki-Sebuguzi, C., Rek, J., Kigozi, S. P., Gonahasa, S., Okiring, J., Nabende, I., Nankabirwa, J. I., Mpeka, B., Kapisi, J., Dorsey, G., and Kamya, M. R. (2022). Assessment of community-wide effects of intermittent preventive treatment for malaria in schoolchildren in Jinja, Uganda (START-IPT trial): a cluster-randomised trial. *Nature Medicine*, 28(5):924–933.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of computational and graphical statistics*, 12(1):156–175.

Vanderslott, S. (2019). Moving from outsider to insider status through metrics: the inclusion of "neglected tropical diseases" into the sustainable development goals. *Journal of Human Development and Capabilities*, 20(4):418–435.

Wienke, A. (2003). Frailty models. *Rostock, Germany Max Planck institute for demographic research*.

Wondaya, S., Kifle, Y. G., Tereda, A. B., and Seyoum, D. (2016). Modeling time to first malaria using spatially correlated conditional autoregressive frailty model.

World Health Organization (2023). Malaria.

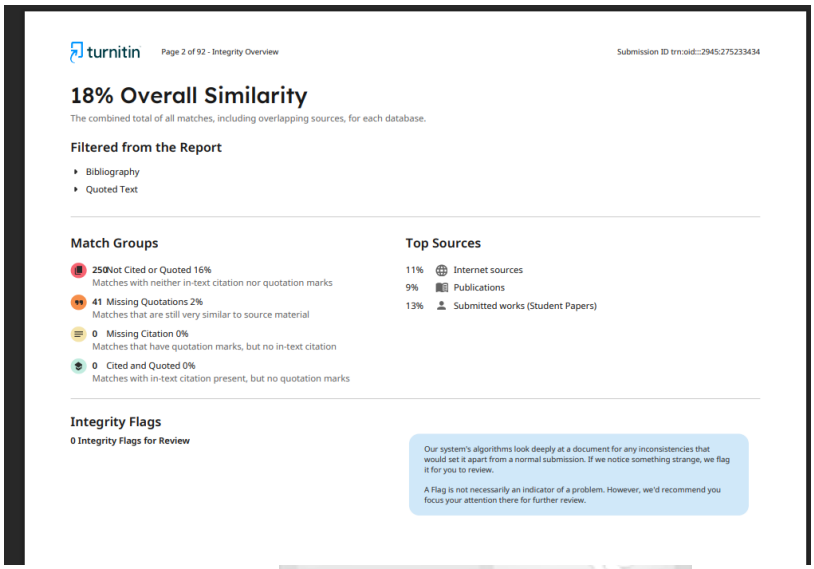
World Health Organization (WHO) (2024). Digital solutions for malaria surveillance. Accessed: 2024-11-25.

Yakum, I. M., Njimanted, G. F., Vukenkeng, A. W., and Nfor, O. N. (2020). Socio-economic determinants of malaria prevention options adoption of households in the north west region of cameroon. *Policy*, 5(2):15–30.

Ye, Y., Patton, E., Kilian, A., Dovey, S., and Eckert, E. (2020). Association between malaria symptoms and prevention behaviors: A hospital-based case-control study in Uganda. *Malaria Journal*, 19(1):1–12.

Appendix A

Similarity index



turnitin Page 2 of 92 - Integrity Overview Submission ID trn:oid:::2945:275233434

18% Overall Similarity

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

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

- 250 Not Cited or Quoted 16%
  Matches with neither in-text citation nor quotation marks
- 41 Missing Quotations 2%
  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

- 11% Internet sources
- 9% Publications
- 13% 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.

Match Groups

- 250 Not Cited or Quoted 16%
  Matches with neither in-text citation nor quotation marks
- 41 Missing Quotations 2%
  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

11% Internet sources
9% Publications
13% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	su-plus.strathmore.edu 3%
2	Internet	eprints.brighton.ac.uk <1%
3	Submitted works	University of Pretoria on 2024-10-14 <1%
4	Internet	core.ac.uk <1%
5	Internet	researchonline.lshtm.ac.uk <1%
6	Internet	s3-eu-west-1.amazonaws.com <1%
7	Publication	John P. Klein, Hans C. van Houwelingen, Joseph G. Ibrahim, Thomas H. Scheike. "H... <1%
8	Submitted works	University of Sunderland on 2024-10-17 <1%
9	Internet	lup.lub.lu.se <1%
10	Internet	monami.hs-mittweida.de <1%

Strathmore UNIVERSITY

28th February 2025

Mr Mwangi Henry,
henry.mwangi@strathmore.edu

Dear Mr Mwangi,

RE: Assessing Latent Factors in Malaria-Related Health Behaviors and Access to Preventive Care using Frailty Models

This is to inform you that SU-ISERC has reviewed and **approved** your above SU-masters proposal. Your application reference number is **SU-ISERC2643/25.** The approval period is from **28th February 2025 to 27th February 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

Appendix B

R code

Here attached is the link to github which contains the dataset and the rcode used in the analysis and in production of the output

https://github.com/Henry4932/dissertation-on-latent-factors-on-malaria-prevention