IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Survival Analysis of Acute Myocardial Infarction Time-to-events Data Using the Kumaraswamy-Logistic Model and Kaplan-Meier Estimation

Ibrahim Abubakar Sadiq[1], Yahaya Zakari[2], Sani Ibrahim Doguwa[3], Mannir Isiya[4], Suleiman Yahaya Suleiman[5], Ayodamola Hephzibah Ajayi[6], Yahaya Yahaya Gambo[7], and Wisdom Samuel[8]

[1,2,3,4,5,6,7,8]Department of Statistics, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria

ARTICLE INFO

Article history:

Received	08 May 2025
Revised	12 May 2005
Accepted	04 July 2025
Available online	06 July 2025

Keywords:

Acute Myocardial Infarction
Survival Analysis
Kumaraswamy-Logistic Model
Kaplan-Meier Estimation
Risk Assessment

ABSTRACT

Background: Acute Myocardial Infarction (AMI) remains a leading cause of morbidity and mortality worldwide. Understanding the influence of patient characteristics on survival outcomes is critical for developing effective clinical interventions.

Objectives: This study aims to evaluate survival outcomes in AMI patients dataset using the Kumaraswamy-logistic survival regression model and the Kaplan-Meier estimator, with a focus on quantifying the effects of covariates such as age, sex, and body mass index (BMI).

Methods: A dataset comprising AMI patient survival times was analyzed. The natural logarithm of survival time was modelled with censoring indicated by d_i (1 = death, 0 = censored). Covariates included age groups (32-51, 52-71, 72-92), sex (male/female), and BMI categories (normal, overweight, obese). Parameter estimation was conducted using Maximum Likelihood Estimation (MLE). Non-parametric survival probabilities were also assessed using the Kaplan-Meier method.

Results: The Kumaraswamy-logistic model revealed significant associations for key predictors. Notably, male sex (β_3 = 1.02299, p = 0.00196) and obesity (β_5 = 1.02584, p = 0.00145) were strongly linked to reduced survival, while advanced age groups also exhibited elevated risk (β_4 = 1.20594, $p < 0.001$ for patients aged 72-92). The scale parameter estimate was σ = 11.85 ($p < 0.0001$) and the shape parameter $\lambda = 4.06$ ($p < 0.0001$), indicating right-skewed survival times with a heavy tail. The Kaplan-Meier survival curve showed high initial survival ($> 80\%$ at 500 days), but a sharp decline over time, with survival probability falling below 50% after approximately 2000 days.

Conclusions: This study demonstrates that survival among AMI patients is significantly influenced by age, sex, and BMI, with older, male, and obese patients experiencing worse outcomes. The Kumaraswamy-logistic model effectively captured the parametric survival structure, while the Kaplan-Meier estimator provided a robust non-parametric benchmark. These results emphasize the need for early risk stratification and tailored interventions in high-risk groups. Future research should validate these findings across broader populations and explore additional prognostic markers to refine survival predictions.

1. Introduction

Acute Myocardial Infarction (AMI), commonly referred to as a heart attack, remains a leading cause of death worldwide despite significant advances in medical treatments and interventions. This condition, caused by the sudden obstruction of blood flow to the heart muscle, leads to substantial mortality and morbidity, particularly in the presence of risk factors such as diabetes, hypertension, and chronic kidney disease. A deeper understanding of survival patterns in AMI patients is crucial for improving clinical outcomes. Achieving this requires the use of robust statistical models to accurately predict

Corresponding author E-mail address: isabubakar@abu.edu.ng
https://doi.org/10.62933/32m8tn21
This work is an open-access article distributed under a CC BY License (Creative Commons Attribution 4.0 International) under https://creativecommons.org/licenses/by-nc-sa/4.0/

survival probabilities and assess the impact of various prognostic factors.

The Kaplan-Meier estimator, a traditional and widely utilized non-parametric approach, is commonly employed to model survival probabilities over time. Its strength lies in its ability to handle censored data, where the exact time of an event (such as death or recovery) is unknown for some patients. This simplicity makes it particularly popular in clinical studies [1, 2, 3, 4, 5]. However, Kaplan-Meier does not accommodate the effects of covariates, limiting its capacity to evaluate the influence of individual risk factors on survival outcomes.

To overcome these limitations, parametric survival models such as the Kumaraswamy logistic survival regression model have gained attention. This model extends the Kumaraswamy distribution, offering greater flexibility to capture diverse survival time distributions [6, 7]. Additionally, it allows for the inclusion of covariates, enabling the identification and quantification of the effects of prognostic factors, such as age, heart function, and comorbid conditions like kidney failure, on patient survival [8]. These features make it a valuable tool for enhancing the predictive accuracy of survival analysis in AMI research [9].

Recent studies highlight the importance of incorporating a wide range of prognostic factors in survival models for AMI patients. For instance, predictive models that include laboratory findings such as troponin, creatinine, and haemoglobin levels alongside clinical variables like hypertension history and previous cardiac events have significantly improved the accuracy of survival predictions [8, 9]. Such advancements allow clinicians to make more informed decisions, ultimately improving post-AMI care and long-term outcomes [8, 9].

In this study, the Kumaraswamy logistic survival regression model will be applied to an AMI dataset, facilitating the simultaneous consideration of multiple covariates. The survival probabilities derived from this model will be compared with those estimated by the Kaplan-Meier method to evaluate their differences and respective strengths. Additionally, residual analysis for both models will provide insights into their accuracy and highlight areas for refinement, ensuring a comprehensive understanding of survival dynamics in AMI patients [10].

By integrating traditional and advanced survival analysis techniques, this research seeks to improve the predictive accuracy of survival outcomes in AMI patients [9]. The findings aim to support the development of personalized treatment strategies, reducing mortality and enhancing recovery outcomes for this vulnerable population [9].

A significant development is the introduction of the Mixed-Effects Parametric Proportional Hazard (MEPPH) model with a generalized log-logistic baseline distribution [2, 3, 4, 5]. This model effectively addresses clustered survival data by incorporating random effects, thereby accounting for unobserved heterogeneity between clusters. The MEPPH model has demonstrated superior performance compared to traditional models, particularly in handling non-monotone hazard rate functions, which are common in medical data, including AMI cases [11].

Additionally, the Kumaraswamy-Log-Logistic distribution has been proposed as a flexible model for survival data, capable of accommodating various hazard rate shapes. This distribution extends the log-logistic model, offering enhanced adaptability in survival analysis [12]. Other contributions to parametric survival distribution include the NOF-G family of distribution by [13], the NGOF-G family of distribution by [14], the NGOF-Et-G family of distribution by [15], the NGOF-OE-G family of distribution by [16], the NETD Using generalized logarithmic function by [17], the extension of T-L distribution by [18], the partial least squares regression by [19], the impact of propensity score-adjusted targeted intervention on survival outcomes among HIV-infected patients by [20], and survival analysis in advanced lung cancer by [5].

The Kaplan-Meier estimator continues to be a fundamental tool in survival analysis, particularly in AMI research. Recent studies

have utilized this method to assess survival probabilities and the impact of various biomarkers on patient outcomes. For instance, elevated admission levels of cystatin C (CysC) in AMI patients have been associated with increased all-cause and cardiovascular mortality over four years, as revealed by Kaplan-Meier survival analysis (Lou *et al*., 2022).

Furthermore, gender differences in AMI prognosis have been examined using Kaplan-Meier survival curves. Findings indicate that male AMI patients may experience different survival outcomes compared to females, underscoring the importance of considering gender in survival analyses [21, 22].

Logistic regression has been increasingly applied in survival analysis to examine the relationship between risk factors and disease events. This approach allows for the assessment of covariate effects on survival probabilities, providing a comprehensive understanding of prognostic factors in AMI patients [23, 24].

In summary, recent literature emphasizes the evolution of survival analysis methodologies, from traditional non-parametric approaches like the Kaplan-Meier estimator to advanced parametric models such as the MEPPH and Kumaraswamy-Log-Logistic distributions. These advancements enhance the accuracy of survival predictions and the understanding of prognostic factors in AMI patients, contributing to improved clinical decision-making and patient outcomes.

1.1 Contribution to the Literature

This study contributes to the survival analysis literature in three significant ways:

Introduction of the Kumaraswamy-Logistic Model in AMI Survival Analysis: This is among the first studies to apply the Kumaraswamy-logistic survival regression model to Acute Myocardial Infarction (AMI) data. While classical models like Weibull or Cox proportional hazards are commonly used, the Kumaraswamy-logistic model offers greater flexibility in capturing skewed hazard behaviors due to its bounded and asymmetric hazard function. This enhances the understanding of long-term survival dynamics, particularly in cases with non-monotonic or non-proportional hazards.

Model-Based Identification of Prognostic Factors:
The study rigorously demonstrates that advanced age, male sex, and obesity are significant predictors of reduced survival in AMI patients, supported by statistically robust maximum likelihood estimates. The model-based approach allows not only inference on survival probabilities but also quantification of the magnitude of risk associated with specific covariates, offering more nuanced clinical interpretations than non-parametric approaches alone.

Integration of Parametric and Non-Parametric Techniques:
By combining the Kumaraswamy-logistic regression model with the Kaplan-Meier estimator, the study highlights the complementary strengths of parametric and empirical methods. This dual approach improves predictive reliability and interpretability, especially when long-term survival estimates are hindered by censoring and small sample sizes.

1.2 Novelty and Relevance

i. This study pioneers the application and evaluation of the Kumaraswamy-logistic model in the clinical context of AMI, where traditional models may fail to fully capture the underlying survival process.
ii. It provides clinically actionable insights, reinforcing the role of obesity
iii. and aging as critical, modifiable risk factors thereby guiding prevention and follow-up strategies.
iv. It contributes a flexible statistical framework that can be extended to other chronic disease datasets where complex hazard shapes and covariate interactions are present.

2. Methodology

This section presents statistical functions and methods that are useful for this research work. These among others include the following: the pdf and CDF of Kumaraswamy-log-logistic distribution, the pdf and survival function of the Kumaraswamy-logistic distribution, the plots of the pdf and survival function of the Kumaraswamy-logistic distribution, and the Kumaraswamy-logistic survival regression model.

2.1 The Kumaraswamy-log-logistic distribution

The researcher [12] defined the cdf and pdf of the Kumaraswamy-log-logistic distribution are respectively given (for $t>0$) and $\alpha>0$ is scale parameter and $\gamma>0$ is shape parameter; and $(\theta>0,\lambda>0)$ extra parameters.

$$f(t)=\frac{\theta\lambda}{\alpha^{\theta\gamma}}t^{\alpha\gamma-1}\left[1+\left(\frac{t}{\alpha}\right)^{\gamma}\right]^{-(\alpha+1)}\left[1-\left(1-\frac{1}{1+\left(\frac{t}{\alpha}\right)^{\gamma}}\right)^{\theta}\right]^{\lambda-1} \tag{1}$$

$$F(t)=1-\left[1-\left(1-\frac{1}{1+\left(\frac{t}{\alpha}\right)^{\gamma}}\right)^{\theta}\right]^{\lambda} \tag{2}$$

The survival and hazard rate functions of the Kumaraswamy-log-logistic distribution are derived as;

$$S(t)=\left[1-\left(1-\frac{1}{1+\left(\frac{t}{\alpha}\right)^{\gamma}}\right)^{\theta}\right]^{\lambda} \tag{3}$$

$$h(t)=\frac{\theta\lambda}{\alpha^{\theta\gamma}}t^{\alpha\gamma-1}\left[1+\left(\frac{t}{\alpha}\right)^{\gamma}\right]^{-(\alpha+1)}\left[1-\left(1-\frac{1}{1+\left(\frac{t}{\alpha}\right)^{\gamma}}\right)^{\theta}\right]^{-1} \tag{4}$$

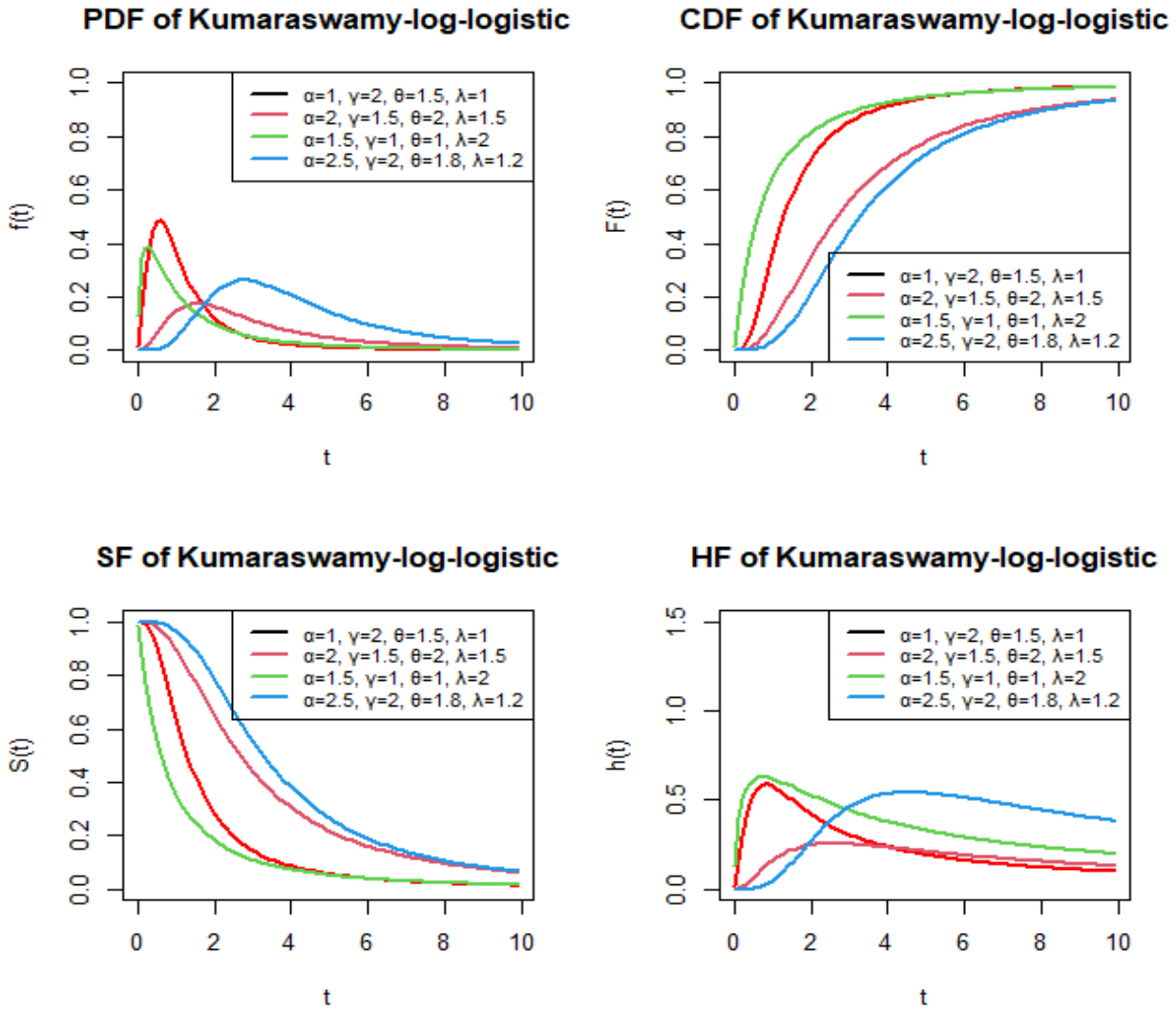

Figure 1: PDF, CDF, SF, and HF Plots of Kumaraswamy-logistic distribution

Figure 1 consists of four plots showing different functions of the Kumaraswamy-log-logistic distribution. Each plot is labelled with the type of function it represents and includes a legend indicating different parameter sets used in the plots. The parameters are (α), (γ), $(\theta), and$ (λ). These plots provide a graphic representation of how the Kumaraswamy-log-logistic distribution behaves under different parameter settings. This can be useful for understanding the distribution's properties and for applications in statistical modelling and data analysis. The PDF shows the likelihood of different values, the CDF shows the cumulative probability, the SF shows the probability of survival beyond a certain time, and the HF shows the instantaneous failure rate.

2.2 The Kumaraswamy-logistic distribution

According to [12], let T be a random variable having the pdf of Kumaraswamy-log-logistic distribution given in equation (1). The random variable $Y = log(T)$ has a Kumaraswamy-logistic density function, parameterized in terms of $\gamma = \sigma^{-1}$ and $\alpha = \exp\{\mu\}$, given by

$$g(y)=\frac{\theta\lambda}{\sigma}\exp\left\{\theta\left(\frac{y-\mu}{\sigma}\right)\right\}\left(1+\exp\left\{\frac{y-\mu}{\sigma}\right\}\right)^{-(\theta+1)}\left[1-1-\left(1-\frac{1}{1+\exp\left\{\frac{y-\mu}{\sigma}\right\}}\right)^{\theta}\right]^{\lambda-1} \quad (5)$$

Where $-\infty<y<\infty$, $-\infty<\mu<\infty$, $\theta>0, \lambda>0, and\ \sigma>0$

The corresponding survival function is given as;

$$S(y)=\left[1-\left(1-\frac{1}{1+\exp\left\{\frac{y-\mu}{\sigma}\right\}}\right)^{\theta}\right]^{\lambda} \quad (6)$$

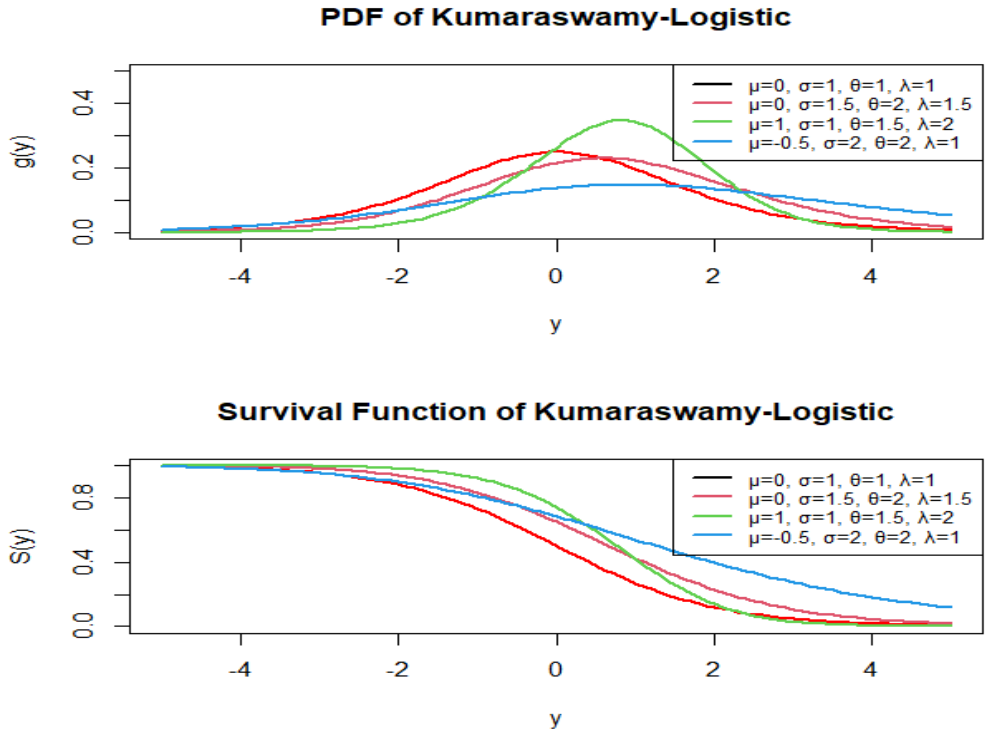

Figure 2: PDF and Survival Function Plots of Kumaraswamy-logistic distribution

Figure 2 contains two plots related to the Kumaraswamy-Logistic distribution. These plots provide a visual representation of how the Kumaraswamy-Logistic distribution behaves under different parameter settings. The PDF plot shows the likelihood of different values of (y), while the Survival Function plot shows the probability of survival beyond a certain value of (y). This information is useful for understanding the distribution's properties and for applications in statistical modeling and data analysis.

The random variable $Z = (Y -)/\sigma$ has a density function

$$d(z)=\theta\lambda\exp\{\theta z\}\left(1+\exp\{z\}\right)^{-(\theta+1)}\left(1-\left[1-\frac{1}{\exp\{z\}}\right]^{\theta}\right)^{\lambda-1} \quad (7)$$

Where $-\infty<z<\infty$

2.3 The Kumaraswamy-logistic Survival Regression Model

Let $X_i = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)^T$ be the explanatory variable vector associated with the ith response variable y_i, for $i, \ldots, n$ [4]. Consider a sample $\left(y_1, v_1\right), \ldots, \left(y_n, v_n\right)$ of n independent observations, where each random response is defined by yi $y_i = \min\left\{\log\left(x_i\right), \log\left(c_i\right)\right\}$, and $\log\left(x_i\right)$ and $\log\left(c_i\right)$ are the log-lifetime and log-censoring, respectively [4]. We consider non-informative censoring such that the observed lifetimes and censoring times are independent [4]. Now, we construct a linear regression model for the response variable y_i based on the Kumaraswamy-logistic distribution given by

$$y_i = X_i^T \beta + \sigma Z_i, \quad i = 1, ..., n \tag{8}$$

where the random error Z_i has the density function in equation (7), $\beta = \left(\beta_1, \beta_2, \ldots, \beta_p\right)^T$; $\sigma > 0$ is a scale parameter, $\theta > 0$ and $\lambda > 0$ are shape parameters and X_i is the vector of explanatory variables modelling the location parameter $\mu_i = X_i^T \beta$. Hence, the location parameter vector $\mu = \left(\mu_1, \mu_2, \ldots, \mu_n\right)^T$ of the Kumaraswamy-logistic model.

The estimated survival function of equation (6) is derived as;

$$S\left(y_i; \hat{\theta}, \hat{\lambda}, \hat{\sigma}\right) = \left[1 - \left(1 - \frac{1}{1 + \exp\left\{\frac{y_i - X_i^T \beta}{\hat{\sigma}}\right\}}\right)^{\hat{\theta}}\right]^{\hat{\lambda}} \tag{9}$$

2.4 Kaplan-Meier Estimator of the Survival Function

The Kaplan-Meier (KM) estimator is a non-parametric method used in survival analysis to estimate the survival function from time-to-event data [3, 4]. It is particularly valuable when dealing with censored data, where the event of interest has not occurred for all subjects within the study period. The Kaplan-Meier estimator also known as the product-limit estimator is the most widely used non-parametric method for estimating the survival function [5]. The Kaplan-Meier estimator provides an estimate of the survival function $S(t)$, which represents the probability that an individual survives beyond time t:

$$S(t) = P(T > t) \tag{10},$$

where T is the random variable representing the time-to-event. Then the K-M estimator of S(t) is defined:

$$\hat{S}(t) = \prod_{t_i \le t}\left(1 - \frac{d_i}{n_i}\right) \tag{11},$$

where t_i: Time of the i^{th} event; d_i: Number of events (e.g., deaths, treatment completion) at t_i; n_i: Number of individuals at risk just before t_i and $\prod$: Product overall event times up to t.

The estimator adjusts for censored data by ensuring that only those individuals still at risk at time t are included in the calculation of the survival probability. It produces a step-function

survival curve, which is widely used in practice.

The researcher [4] showed the consistency of this estimator. They obtained the approximate formula for the variance of the estimator,

$$Var\left(\hat{S}_n(t)\right)=\left[\hat{S}_n(t)\right]^2\sum_{\{i:T_{(i)}\le t\}}\frac{\sigma_i}{(n-i)(n-i+1)} \quad (12)$$

When there are no ties, but when ties occur,

$$Var\left(\hat{S}_n(t)\right)=\left[\hat{S}_n(t)\right]^2\sum_{\{i:T_{(i)}\le t\}}\frac{\sigma_i}{(n_i-1)n_i} \quad (13)$$

2.4.1 Confidence interval for S(t)

We first estimate the C.I for the unknown survival function S(t) for a fixed value of t. we know that when the sample is large, the standardized version [4],

$$\frac{\hat{S}(t)-S(t)}{\sqrt{Var\left(\hat{S}(t)\right)}} \sim N(0,1) \quad (14)$$

Then for a given t, this would lead to asymptotic (1-α) percent C.I for S(t): Where $Z_{1-\frac{\alpha}{2}}$ is the $(1-\alpha)100\%$ upper critical point of the standard normal distribution [4].

3. Results and discussion

In this section, we present the results and discussion of the Kumaraswamy-logistic survival regression model and Kaplan-Meier probability with applications to the acute myocardial infarction (AMI) dataset [9].

3.1 The AMI Dataset and its Exposure Variables

The Worcester Heart Attack Study provides a rich dataset on AMI patients [9], offering valuable insights into survival trends, risk factors, and treatment outcomes.

Table 1: Description of the AMI Dataset and its Exposure Variables

Variables	Descriptions
y (Log Survival Time)	The natural logarithm of the observed survival time (in days).
d (Censoring)	Indicates whether the patient was censored (0) or died (1).
Age	Categories patients based on age groups: 32-51 (early to mid-adulthood), 52-71 (late adulthood), 72-92 (elderly).
Sex	Indicates the gender of the patient (Male, Female).
BMI	Categories patients based on body mass index (BMI): 18.5-24.9 (normal), 25-29.9 (overweight), >= 30 (obese).

The response variable**:** y: Natural logarithm of observed survival time (in days); censoring variable: d: 0 (alive at study end or lost to follow-up), 1 (death due to AMI); exposure variables: age groups: $x1, x2, x3$; sex: $x4, x5$ and the BMI categories: $x6, x7, x8, x9$.

Table 2: MLE of parameters of the Kumaraswamy-logistic survival regression model on the AMI Dataset

Coefficients	Estimates	Standard Error	Z value	Pr(z)
Λ	4.05585	0.8649	4.6894	2.74e-06
θ	266.19635	270.6229	0.9836	0.325291
σ	11.85107	2.37254	4.9951	5.88e-07
β_0	1.12889	0.36472	3.0952	0.001967
β_1	0.62314	0.42577	1.4636	0.14331
β_2	0.44694	0.3312	1.3495	0.177186
β_3	1.02299	0.33041	3.0961	0.001961
β_4	1.20594	0.29202	4.1297	3.63e-05
β_5	1.02584	0.32206	3.1852	0.001447
β_6	-1.6115	0.89559	-1.7994	0.07196
β_7	1.63916	0.40682	4.0292	5.60e-05
β_8	1.90134	0.402	4.7297	2.25e-06
β_9	2.25803	0.53881	4.1908	2.78e-05

Table 1 presents the results of the Kumaraswamy-logistic survival regression model effectively captures the relationship between survival time and key predictors in AMI patients. The scale parameters λ ($p < 0.001$) *and* σ ($p < 0.001$) significantly influence the survival distribution, highlighting their crucial role in modeling survival time. However, the shape parameter θ is not significant $(\mathrm{p} = 0.325)$, suggesting that the effect of this parameter on the survival curve is less pronounced. The oldest age group $(\mathrm{x}_3, 72-92$ years) significantly increases survival time $(\mathrm{p} = 0.002)$, suggesting that elderly patients exhibit better survival outcomes compared to younger groups. Both male $(\beta_4,\ p < 0.001)$ and female $(\beta_5, p = 0.001)$ patients demonstrate significantly improved survival times, with males showing a slightly higher effect than females. Overweight $(\mathrm{x}_7, \mathrm{p} < 0.001)$ and obese $(x_8, \mathrm{p} < 0.001)$ categories positively impact survival, indicating that higher BMI may provide a protective effect. Conversely, normal BMI (x_6) shows a marginally negative effect $(\mathrm{p} < 0.0072)$. This β_9 is highly significant $(\mathrm{p} < 0.001)$ and positively influences survival. These results suggest that demographic (age, gender) and clinical factors (BMI) significantly influence survival outcomes in AMI patients. The protective role of higher BMI categories aligns with the "obesity paradox" observed in cardiovascular disease studies. The significant scale parameters indicate that the Kumaraswamy-logistic model appropriately captures survival time trends, though additional exploration of the shape parameter may refine its application. These findings provide valuable insights for personalized treatment strategies and risk stratification among AMI patients.

3.2 Kaplan-Meier Survival Probability Analysis

The Kaplan-Meier survival analysis results in Tables 3 to 11 provide valuable insights into the survival probability of patients with Acute Myocardial Infarction (AMI) over time.

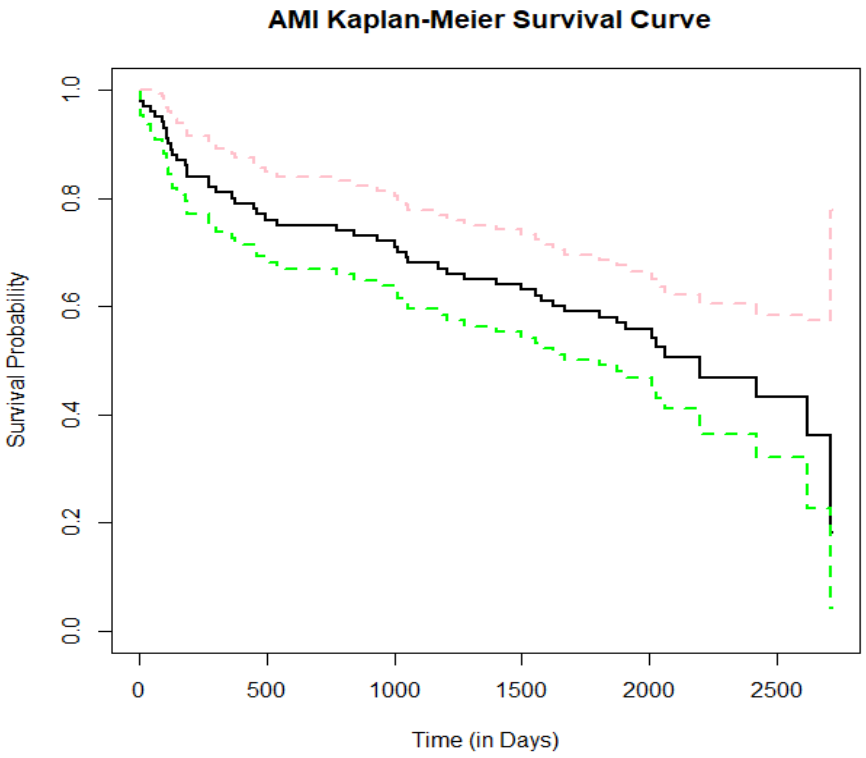

Figure 3: plot of the Kaplan-Meier Survival curve

Figure 3 presents the Kaplan-Meier survival curve for patients with Acute Myocardial Infarction (AMI). The estimated survival probability decreases steadily over time, with a more rapid decline during the early follow-up period (first 1000 days). The survival probability falls below 50% after approximately 2000 days. The widening of the 95% confidence intervals toward the right of the curve reflects increasing uncertainty due to censoring and fewer individuals at risk. These findings highlight the need for early post-AMI monitoring and long-term management strategies to improve survival outcomes. The curve confirms that mortality is front-loaded, with a significant number of events occurring relatively early. The findings support more intensive monitoring and intervention in the early phase post-AMI.

Table 3: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x1 = 1, x2 = 0, x3 = 0

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
1048	16	1	0.938	0.0605	0.826	1
1172	15	1	0.875	0.0827	0.727	1
1401	14	1	0.812	0.0976	0.642	1
1577	13	1	0.75	0.1083	0.565	0.995

Table 4: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x1 = 0, x2 = 1, x3 = 0

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
6	35	1	0.971	0.0282	0.9178	1

14	34	1	0.943	0.0392	0.869	1
44	33	1	0.914	0.0473	0.8261	1
128	32	1	0.886	0.0538	0.7863	0.998
182	31	1	0.857	0.0591	0.7487	0.981
302	30	1	0.829	0.0637	0.7127	0.963
538	29	1	0.8	0.0676	0.6779	0.944
774	28	1	0.771	0.071	0.6441	0.924
1002	27	1	0.743	0.0739	0.6113	0.903
1011	26	1	0.714	0.0764	0.5793	0.881
1278	25	1	0.686	0.0785	0.5479	0.858
1669	24	1	0.657	0.0802	0.5173	0.835
2624	2	1	0.329	0.2358	0.0805	1

Table 5: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x1 = 0, x2 = 0, x3 = 1

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
6	49	1	0.98	0.0202	0.9408	1
62	48	1	0.959	0.0283	0.9054	1
89	47	1	0.939	0.0342	0.874	1
98	46	1	0.918	0.0391	0.8448	0.998
104	45	1	0.898	0.0432	0.8171	0.987
107	44	1	0.878	0.0468	0.7904	0.974
114	43	1	0.857	0.05	0.7646	0.961
123	42	1	0.837	0.0528	0.7394	0.947
148	41	1	0.816	0.0553	0.7148	0.932
187	40	1	0.796	0.0576	0.6907	0.917
189	39	1	0.776	0.0596	0.6671	0.902
274	38	2	0.735	0.0631	0.6209	0.869
363	36	1	0.714	0.0645	0.5984	0.853
374	35	1	0.694	0.0658	0.5761	0.836
451	34	1	0.673	0.067	0.5542	0.818
461	33	1	0.653	0.068	0.5325	0.801
492	32	1	0.633	0.0689	0.5111	0.783
841	31	1	0.612	0.0696	0.49	0.765
936	30	1	0.592	0.0702	0.469	0.747
1054	29	1	0.571	0.0707	0.4484	0.728
1205	28	1	0.551	0.0711	0.428	0.709
1497	27	1	0.531	0.0713	0.4078	0.69
1557	26	1	0.51	0.0714	0.3878	0.671
1624	25	1	0.49	0.0714	0.3681	0.652
1806	24	1	0.469	0.0713	0.3485	0.632
1874	20	1	0.446	0.0715	0.3257	0.611
1907	19	1	0.422	0.0715	0.3032	0.589
2012	12	1	0.387	0.0737	0.2667	0.562
2031	11	1	0.352	0.0749	0.232	0.534
2065	10	1	0.317	0.0752	0.1989	0.505

2201	4	1	0.238	0.0888	0.1142	0.494
2421	3	1	0.158	0.0877	0.0535	0.469
2710	1	1	0.000	0.0000	0.0000	0.000

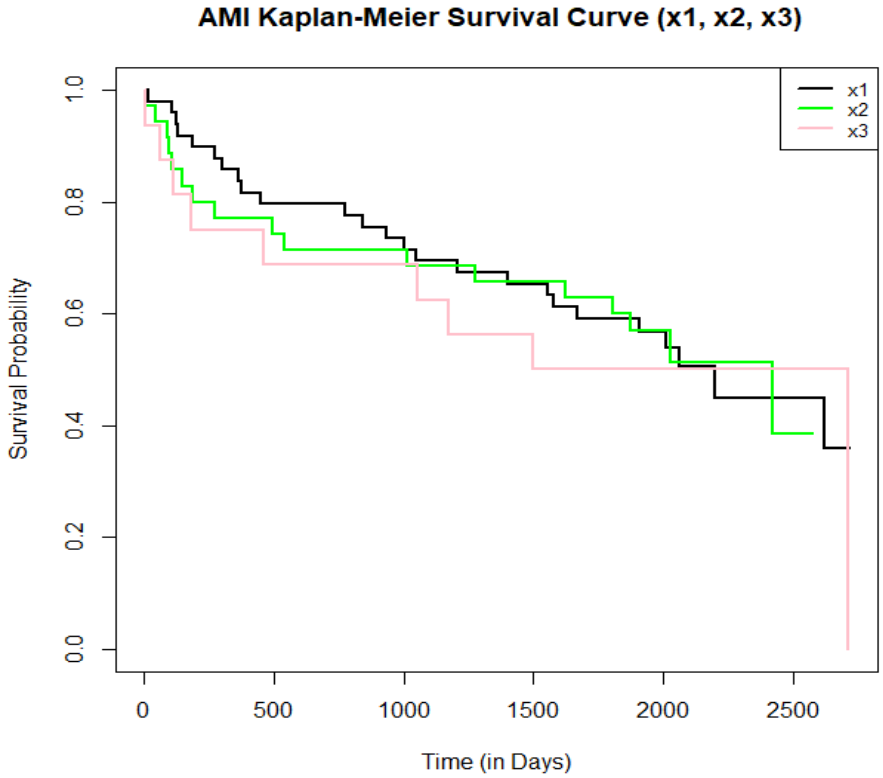

Figure 4: plot of the Kaplan-Meier Survival curve

Table 6: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x4 = 1, x5 = 0

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
14	35	1	0.971	0.0282	0.918	1
62	34	1	0.943	0.0392	0.869	1
89	33	1	0.914	0.0473	0.826	1
98	32	1	0.886	0.0538	0.786	0.998
104	31	1	0.857	0.0591	0.749	0.981
128	30	1	0.829	0.0637	0.713	0.963
148	29	1	0.8	0.0676	0.678	0.944
187	28	1	0.771	0.071	0.644	0.924
302	27	1	0.743	0.0739	0.611	0.903
363	26	1	0.714	0.0764	0.579	0.881
374	25	1	0.686	0.0785	0.548	0.858
461	24	1	0.657	0.0802	0.517	0.835
841	23	1	0.629	0.0817	0.487	0.811
1002	22	1	0.6	0.0828	0.458	0.786
1011	21	1	0.571	0.0836	0.429	0.761
1172	20	1	0.543	0.0842	0.401	0.736
1577	19	1	0.514	0.0845	0.373	0.71

1806	18	1	0.486	0.0845	0.345	0.683
1874	14	1	0.451	0.0853	0.311	0.653
2031	10	1	0.406	0.0879	0.266	0.62
2065	9	1	0.361	0.0889	0.223	0.585
2201	6	1	0.301	0.0922	0.165	0.549
2710	1	1	0	0.0000	0.000	0.000

Table 7: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x4 = 0, x5 = 1

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
6	65	2	0.969	0.0214	0.928	1
44	63	1	0.954	0.026	0.904	1
107	62	1	0.938	0.0298	0.882	0.999
114	61	1	0.923	0.0331	0.861	0.99
123	60	1	0.908	0.0359	0.84	0.981
182	59	1	0.892	0.0384	0.82	0.971
189	58	1	0.877	0.0407	0.801	0.961
274	57	2	0.846	0.0448	0.763	0.939
451	55	1	0.831	0.0465	0.744	0.927
492	54	1	0.815	0.0481	0.726	0.915
538	53	1	0.8	0.0496	0.708	0.903
774	52	1	0.785	0.051	0.691	0.891
936	51	1	0.769	0.0523	0.673	0.879
1048	50	1	0.754	0.0534	0.656	0.866
1054	49	1	0.738	0.0545	0.639	0.853
1205	48	1	0.723	0.0555	0.622	0.84
1278	47	1	0.708	0.0564	0.605	0.827
1401	46	1	0.692	0.0572	0.589	0.814
1497	45	1	0.677	0.058	0.572	0.801
1557	44	1	0.662	0.0587	0.556	0.787
1624	43	1	0.646	0.0593	0.54	0.774
1669	42	1	0.631	0.0599	0.524	0.76
1907	35	1	0.613	0.0608	0.504	0.744
2012	24	1	0.587	0.0634	0.475	0.726
2421	8	1	0.514	0.0883	0.367	0.72
2624	4	1	0.385	0.1295	0.199	0.744

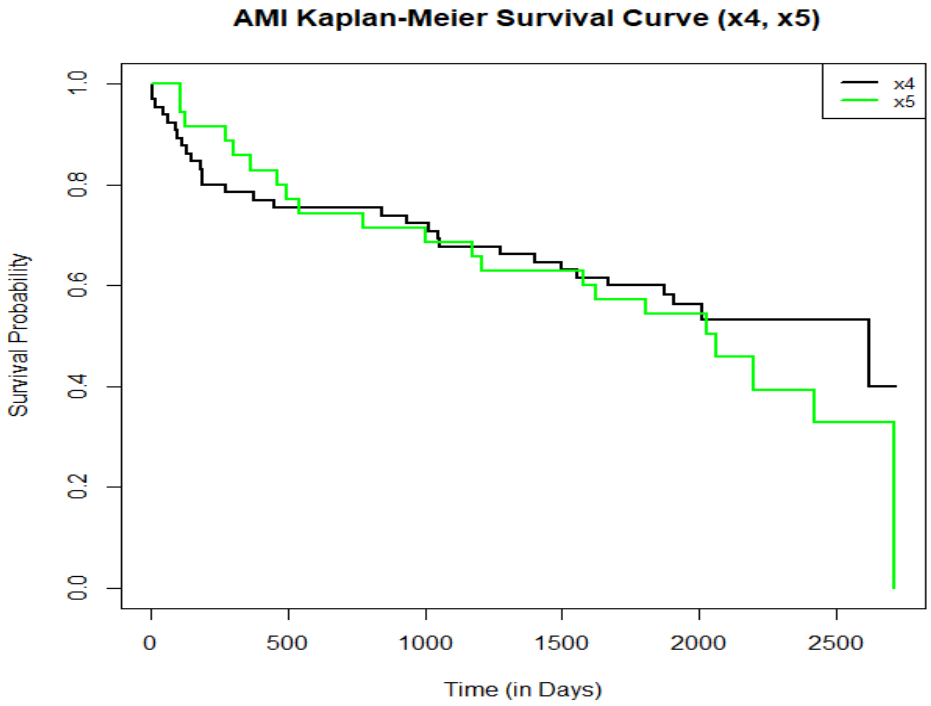

Figure 5: plot of the Kaplan-Meier Survival curve

Table 8: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x6 = 1, x7 = 0, x8 = 0, x9 = 0

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
62	3	1	0.667	0.272	0.2995	1
123	2	1	0.333	0.272	0.0673	1
187	1	1	0	0.000	0.0000	0.00

Table 9: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x6 = 0, x7 = 1, x8 = 0, x9 = 0

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
14	33	1	0.97	0.0298	0.913	1
44	32	1	0.939	0.0415	0.861	1
89	31	1	0.909	0.05	0.816	1
98	30	1	0.879	0.0568	0.774	0.998
107	29	1	0.848	0.0624	0.735	0.98
128	28	1	0.818	0.0671	0.697	0.961
148	27	1	0.788	0.0712	0.66	0.94
182	26	1	0.758	0.0746	0.625	0.919
274	25	1	0.727	0.0775	0.59	0.896
374	24	1	0.697	0.08	0.557	0.873
451	23	1	0.667	0.0821	0.524	0.849
461	22	1	0.636	0.0837	0.492	0.824
492	21	1	0.606	0.0851	0.46	0.798
538	20	1	0.576	0.086	0.43	0.772
841	19	1	0.545	0.0867	0.399	0.745

1278	18	1	0.515	0.087	0.37	0.717
1401	17	1	0.485	0.087	0.341	0.689
1497	16	1	0.455	0.0867	0.313	0.661
1557	15	1	0.424	0.086	0.285	0.631
1806	14	1	0.394	0.0851	0.258	0.601
1874	13	1	0.364	0.0837	0.232	0.571
2012	8	1	0.318	0.0847	0.189	0.536

Table 10: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x6 = 0, x7 = 0, x8 = 1, x9 = 0

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
6	35	1	0.971	0.0282	0.918	1
189	34	1	0.943	0.0392	0.869	1
274	33	1	0.914	0.0473	0.826	1
302	32	1	0.886	0.0538	0.786	0.998
363	31	1	0.857	0.0591	0.749	0.981
774	30	1	0.829	0.0637	0.713	0.963
936	29	1	0.8	0.0676	0.678	0.944
1011	28	1	0.771	0.071	0.644	0.924
1172	27	1	0.743	0.0739	0.611	0.903
1907	20	1	0.706	0.079	0.567	0.879
2065	15	1	0.659	0.0866	0.509	0.852
2201	8	1	0.576	0.108	0.399	0.832
2421	7	1	0.494	0.1199	0.307	0.795
2710	1	1	0	0.0000	0.000	0.000

Table 11: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

x6 = 0, x7 = 0, x8 = 0, x9 = 1

Survival Time	Number at Risk	Number of Eevent	Survival Probability	Standard Error	95% Lower CI	95% Upper CI
6	29	1	0.966	0.0339	0.9013	1
104	28	1	0.931	0.0471	0.8432	1
114	27	1	0.897	0.0566	0.7923	1
1002	26	1	0.862	0.064	0.7453	0.997
1048	25	1	0.828	0.0701	0.7009	0.977
1054	24	1	0.793	0.0752	0.6586	0.955
1205	23	1	0.759	0.0795	0.6178	0.932
1577	22	1	0.724	0.083	0.5784	0.907
1624	21	1	0.69	0.0859	0.5403	0.88
1669	20	1	0.655	0.0883	0.5031	0.853
2031	10	1	0.59	0.1009	0.4217	0.825
2624	2	1	0.295	0.2145	0.0708	1

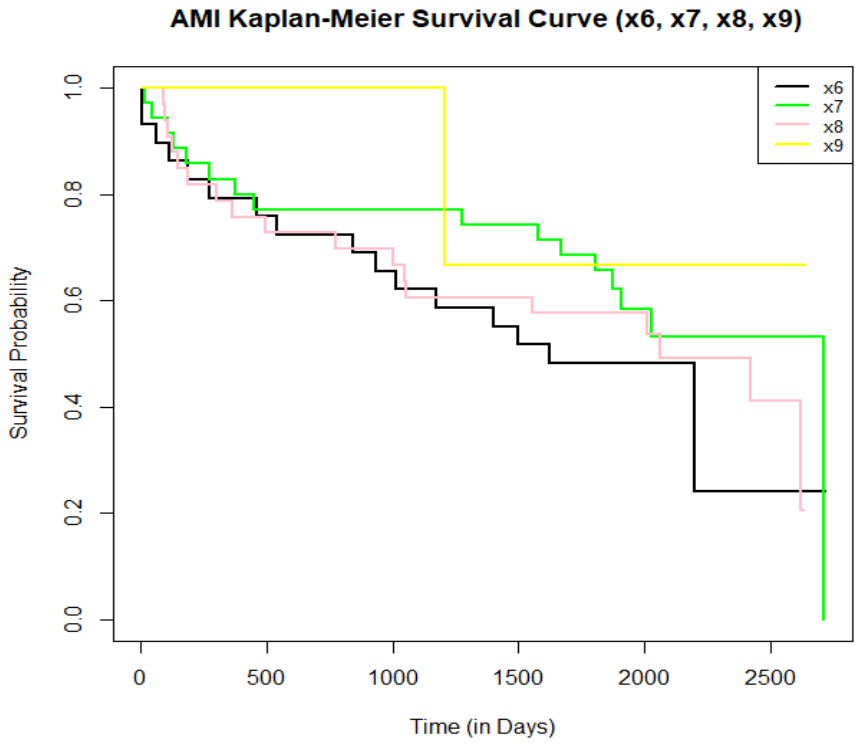

Figure 6: plot of the Kaplan-Meier Survival curve

The "Number at Risk" column indicates how many patients are still being observed (i.e., are at risk of an event) at each time point. For instance, at time 6, there are 49 patients at risk, and this number decreases as time progresses due to events (deaths) or censored data.

The "Number of Events" shows how many patients experienced the event of interest (death) at each specific time point. For example, there was one event at times 6, 62, and 89, indicating that one patient died at each of those times. At time 274, two patients experienced the event.

The "Survival Probability" column provides the estimated probability that a patient will survive beyond each time point. The survival probability decreases over time as more patients experience the event. At time 6, the survival probability is 0.98, indicating a 98% chance of survival at that time. By time 2421, the survival probability has dropped to 0.158, showing a much lower chance of survival as time progresses.

The "Standard Error" represents the uncertainty in the survival probability estimate at each time point. A lower standard error (e.g., 0.0202 at time 6) indicates more precision in the estimate, while a higher standard error (e.g., 0.0888 at time 2201) suggests greater uncertainty. The standard error increases as time progresses due to a decreasing sample size (i.e., fewer patients at risk).

The confidence intervals provide a range within which we can be 95% confident that the true survival probability lies. At time 6, the survival probability is 0.98, with a 95% confidence interval of 0.9408 to 1, meaning that the true survival probability is likely within this range. As time progresses, the confidence intervals widen, reflecting increasing uncertainty. By time 2421, the survival probability is 0.158, with a wider confidence interval of 0.0535 to 0.469.

In the early stages of the study (e.g., up to time 187), survival probabilities are relatively high, indicating that the majority of patients survive these early time points. The survival probability only drops to 0.796 by time 187, showing a fairly strong survival rate in the short term for AMI patients. After time 2000, the survival probability drops significantly. By the time 2012, the survival

probability was 0.387, and by time 2421, it had decreased to 0.158, indicating that long-term survival is low for this cohort of AMI patients. As time progresses, the number of patients at risk decreases, leading to wider confidence intervals and higher standard errors. For example, at time 2201, the standard error is 0.0888, and the 95% confidence interval is very wide (from 0.1142 to 0.494), indicating less precision in estimating survival probabilities for later time points. This widening reflects the smaller sample size and greater variability in survival times at later stages. The last row (time 2710) shows a survival probability of 0 with missing standard error and confidence intervals, suggesting that the last patient died at this time point, leaving no one else at risk.

The Kaplan-Meier analysis reveals a steady decline in survival probability over time, with early survival being strong but significantly decreasing after around 1000 days. This analysis provides crucial insights into the risk factors and survival outcomes for AMI patients and suggests that survival diminishes notably as time progresses, especially in the long term.

4. Conclusions

This study presents a comprehensive survival analysis of Acute Myocardial Infarction (AMI) patients' time-to-events data using the Kumaraswamy-logistic survival regression model alongside the non-parametric Kaplan-Meier estimator. The results affirm that patient-specific covariates particularly age, sex, and body mass index (BMI) significantly influence survival outcomes.

The Kumaraswamy-logistic model successfully captured the parametric behaviour of survival time, revealing that advanced age ($\beta_4 = 1.20594, p < 0.001$), male sex ($\beta_3 = 1.02299, p = 0.00196$), and obesity ($\beta_5 = 1.02584, p = 0.00145$) are strong predictors of decreased survival. The significant shape ($\lambda = 4.06, p < 0.0001$) and scale ($\sigma = 11.85,\ p < 0.0001$) parameters indicate a right-skewed distribution with decreasing hazard over time. Complementarily, the Kaplan-Meier curve revealed high initial survival probabilities that declined notably after approximately 2000 days, with survival probability dropping below 50%.

Clinically, these findings call attention to the urgent need for personalized risk stratification and intervention strategies, especially targeting elderly, male, and obese patients. The combination of advanced parametric and non-parametric methods enhances the robustness of survival inference, supporting evidence-based improvements in post-AMI management and long-term care.

Future research should validate these findings using larger and more diverse cohorts and explore additional prognostic variables to refine predictive accuracy and clinical applicability.

References

[1] Xiao, J., Mo, M., Wang, Z., Zhou, C., Shen, J., Yuan, J., ... & Zheng, Y. (2022). The application and comparison of machine learning models for the prediction of breast cancer prognosis: retrospective cohort study. *JMIR medical informatics*, *10*(2), e33440.

[2] Usman , . A. ., Doguwa, S. I. ., Sadiq, I. A. ., & Akor, A. (2025). Exploring Accelerated Failure Time Models for Tuberculosis Survival: Loglogistic and Weibull Survival Regression Model. *Journal of Science Research and Reviews*, *2*(1), 27-36. https://doi.org/10.70882/josrar.2025.v2i1.27

[3] Akor, A., Sadiq, I. A., Usman, A., Doguwa, S. I., & Akor, L. O. (2025). Advanced Survival Modeling of Tuberculosis Patients: Insights from Exponential and Weibull AFT Models. *Fudma Journal of Sciences*, *9*(3), 169 - 182. https://doi.org/10.33003/fjs-2025-0903-3350

[4] Semary, H., Sadiq, I. A., Doguwa, S. I. S., Ishaq, A. I., Suleiman, A. A., Daud, H., & Abd Elgawad, M. A. (2025). Advancing survival regression using the NGOF exponentiated Weibull distribution for vesicovaginal fistula and radiation data applications. *Journal of Radiation Research and Applied Sciences*, *18*(2), 101497.

[5] Sadiq, I. A., Yahaya, G. Y., Hephzibah, A. A., Yahaya, S. S., Wisdom, S., Isiya, M., & Bello, A. (2025). Survival Analysis in Advanced Lung

Cancer: A Weibull Survival Regression Model, *in Proceedings of the 1st International Online Conference on Clinical Reports*, 19–20 March 2025, MDPI: Basel, Switzerland.

[6] Usman, A., Yusuf, H., Yahaya, A., Sadiq, I. A., Akanji, O. B., & Aliyu, S. A. (2025). Comparing the Methods of Estimators of the Modified Inverted Kumaraswamy Distribution Using Inverse Power Function. *UMYU Scientifica*, *4*(1), 325–335. https://doi.org/10.56919/usci.2541.032

[7] Yusuf, H., Usman, A., Yahaya, A., Sadiq, I. A., Bello, O. A., & Adamu, S. A. (2025). Modified Inverted Kumaraswamy Distribution Using Inverse Power Function: Properties and Applications. *Fudma Journal of Sciences*, *9*(1), 234 - 239. https://doi.org/10.33003/fjs-2025-0901-3177

[8] de Oliveira, R. P., de Oliveira Peres, M. V., Achcar, J. A., & Martinez, E. Z. (2023). A new class of bivariate Sushila distributions in the presence of right-censored and cure fractions. *Brazilian Journal of Probability and Statistics*, *37*(1), 55-72.

[9] Mercado-Lubo, R., Yarzebski, J., Lessard, D., Gore, J., & Goldberg, R. J. (2020). Changing trends in the landscape of patients hospitalized with acute myocardial infarction (2001 to 2011)(from the Worcester Heart Attack Study). *The American journal of cardiology*, *125*(5), 673-677.

[10] Gao, N., Qi, X., Dang, Y., Li, Y., Wang, G., Liu, X., ... & Fu, J. (2020). Establishment and validation of a risk model for prediction of in-hospital mortality in patients with acute ST-elevation myocardial infarction after primary PCI. *BMC Cardiovascular Disorders*, *20*, 1-10.

[11] Peter, M. W., Mwalili, S. M., Wanjoya, A. K., & Muse, A. H. (2023). Mixed-Effects Parametric Proportional Hazard Model with Generalized Log-Logistic Baseline Distribution. *Journal of Data Analysis and Information Processing*, *11*(2), 81-102.

[12] De Santana, T. V. F., Ortega, E. M., Cordeiro, G. M., & Silva, G. O. (2012). The Kumaraswamy-log-logistic distribution. *Journal of Statistical Theory and Applications*, *11*(3), 265-291.

[13] Sadiq, I. A., Doguwa, S. I., Yahaya, A., & Garba, J. (2022). New Odd Frechet-G Family of Distribution with Statistical Properties and Applications. *AFIT Journal of Science and Engineering Research*, 2(2), 84-103.

[14] Sadiq, I. A., Doguwa, S. I., Yahaya, A., & Garba, J. (2023a). New Generalized Odd Frechet-G (NGOF-G) Family of Distribution with Statistical Properties and Applications. *UMYU Scientifica*, *2*(3), 100-107.

[15] Sadiq, I. A., Doguwa, S. I. S., Yahaya, A., & Usman, A. (2023b). Development of New Generalized Odd Fréchet-Exponentiated-G Family of Distribution. *UMYU Scientifica*, *2*(4), 169-178.

[16] Sadiq, I. A., Doguwa, S. I. S., Yahaya, A., & Garba, J. (2023c). New Generalized Odd Fréchet-Odd Exponential-G Family of Distribution with Statistical Properties and Applications. *FUDMA Journal of Sciences*, *7*(6), 41-51.

[17] Obafemi, A. A., Usman, A., Sadiq, I. A., & Okon, U. (2024). A New Extension of Topp-Leone Distribution (NETD) Using Generalized Logarithmic Function. *UMYU Scientifica*, *3*(4), 127-133.

[18] Habu, L., Usman, A., Sadiq, I. A., & Abdullahi, U. A. (2024). Estimation of Extension of Topp-Leone Distribution using Two Different Methods: Maximum Product Spacing and Maximum Likelihood Estimate. *UMYU Scientifica*, *3*(2), 133-138.

[19] Sadiq, I. A. (2025, April). A Study on the Performance of the Partial Least Squares Regression in Handling Multicollinearity Using Simulated Data. In *Royal Statistical Society Nigeria Local Group Annual Conference Proceedings* (pp. 318-328).

[20] Sadiq, I. A., Usman, A., Kajuru, J. Y., Zakari, Y., Ishaq, A. I., & Sadiq, I. Z. (2025). Impact of Propensity Score-Adjusted Targeted Intervention on Survival Outcomes Among HIV-Infected Patients: A Clinical Trial Analysis, *in Proceedings of the 1st International Online Conference on Clinical Reports*, 19–20 March 2025, MDPI: Basel, Switzerland.

[21] Wang, P., Yao, J., Xie, Y., & Luo, M. (2020). Gender-specific predictive markers of poor prognosis for patients with acute myocardial infarction during a 6-month follow-up. *Journal of Cardiovascular Translational Research*, *13*, 27-38.

[22] Sadiq, I. A., & Komali, K. (2020). On the Study of Clear Causal Risk Factors of Diabetes MellitusUsing Multiple Regression. *Asian Journal of Probability and Statistics*, *9*(4), 29-47.

[23] Sadiq, I. A., Raghav, J. S., & Sharma, S. K. (2020). Conglomeration of General Linear Model for Epilepsy Clinical Neuroimaging. *Asian Journal of Probability and Statistics*, *10*(2), 1-12.

[24] Abbott, R. D. (1985). Logistic regression in survival analysis. *American journal of epidemiology*, *121*(3), 465-471.