Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics

AN APPLICATION OF THE KAPLAN-MEIER SURVIVABILITY PROBABILITY AND A LOG-EXTENDED WEIBULL SURVIVAL REGRESSION MODEL TO THE ACUTE MYOCARDIAL INFARCTION DATA SET

Hauwa'u Yusuf[1,*], Ibrahim Abubakar Sadiq[1], Ajayi Ayodamola Hephzibah[1], Abubakar Usman[1], Jamila Abdullahi[1], Reuben Oluwabukunmi David[1], Saudat Ali Adamu[1] and Hafsat Abdulkarim[1]

[1]*Department of Statistics, Ahmadu Bello University, Zaria*

*Corresponding Author: hauwau0801691@gmail.com

ABSTRACT

Acute Myocardial Infarction (AMI) remains a leading cause of mortality worldwide, necessitating a thorough understanding of survival outcomes and risk factors associated with the condition. This study applies the Log-Extended Weibull Survival Regression Model and the Kaplan-Meier Estimator to examine survival probabilities and identify key prognostic factors among AMI patients. The study utilized survival data from patients diagnosed with AMI. The Log-Extended Weibull model was employed to assess the impact of covariates, while the Kaplan-Meier estimator provided non-parametric survival probability estimates over time. Model residuals were analyzed to evaluate the overall fit. Maximum likelihood estimates (MLE) of the model parameters were derived, and survival probability curves from both methods were compared. The Log-Extended Weibull model identified several key covariates as significant predictors of survival, including β_2 (HR = 1.447, p < 2.20e-16), β_3 (HR = 4.851, p < 2.20e-16), and β_7 (HR = -2.638, p < 2.20e-16), highlighting the critical role of these factors in AMI patient outcomes. The Kaplan-Meier survival curve indicated that 98% of patients survived beyond 6 days, but survival declined gradually to 18% by day 2710. Comparison of the survival curves demonstrated consistency between the two methods, though the regression model provided deeper insights into the effect of covariates on survival. This study demonstrates the effectiveness of both the Log-Extended Weibull model and the Kaplan-Meier estimator in analyzing AMI survival. Significant covariates were identified, providing valuable insights into patient prognosis. The findings underscore the importance of early intervention, long-term patient management, and targeted care strategies to improve survival outcomes. This research contributes to the understanding of survival patterns in AMI patients and offers evidence for better clinical decision-making and personalized treatment approaches.

Keywords: Acute Myocardial Infarction, Log-Extended Weibull Model, Kaplan-Meier Estimator, Survival Analysis, Prognostic Factors, Risk Assessment.

1. INTRODUCTION

The Weibull Distribution also known as the lifetime distribution is a continuous probability distribution used to analyse life data, model failure times and assess product reliability (Silva *et al.*, 2009). The Weibull distribution is also widely used in reliability as a model for time to failure. It can also be used to fit a large range of data from other fields such as economics, biology, and engineering sciences. It is an extreme probability distribution value frequently used to model reliability, survival and other data (Silva *et al.*, 2009). The Weibull distribution is mostly used because of its flexibility, it can be used to simulate various distributions like the normal and exponential distributions. The Weibull distribution is mostly used in reliability analysis and life data analysis because of its ability to adapt to different situations. The probability density function usually describes the distribution function, parameters in the distribution determine the shape, scale and location of the probability density function. The reliability of the Weibull distribution can be measured with the help of parameters.

The Weibull distribution has been modelled over time to be able to adapt to more practical problems and one of these is the bathtub-shaped data distribution which is mostly found in reliability engineering and survival analysis. The Weibull distribution, log-logistic and log-normal distribution are suitable only where the failure rate is constant, monotone or unimodal, that is why the log-extended Weibull distribution has been adopted to be used when these conditions do not apply to the distribution of the data.

Coderio *et al.* (2021) The induction of one or more parameter(s) in parent distributions opened new doors for flexible modelling in modern distribution theory. Among well-established generalized (G) classes for flexible modelling, the exponentiated-G, Marshall-Olkin-G and odd log-logistic-G families offer induction of one additional parameter while the beta-G and Kumaraswamy-G classes offer two extra shape parameters. The Marshall-Olkin-odd-loglogistic-G (MOOLL-G) family serves as an alternative to the beta-G and Kumaraswamy-G classes. A new motivation for the MOOLL-G family for competing risk scenarios, some useful properties, and parameter estimation are addressed. The new log-MOOLL-Weibull regression is useful for the analysis of real-life data. The accuracy of the estimates and the residuals is addressed via Monte Carlo simulations. The presented models outperform some other well-known models.

The Kaplan-Meier method shows statistics which gives the researcher another way to present survival data. This method provides a survival graph over time which shows the estimated percent survival of the subject(s) at each point in time. It is a method of summarizing survival data, which uses all the cases in a series not just those followed up until the selected cut-off. Using this method we divide the follow-up data into many small intervals determining for each interval the number of cases followed up over that interval and the number of events of interest, (for example death or failure) during each period. When the surviving proportion is multiplied by the surviving proportions for each of the preceding periods, a probability is obtained which is called the survival probability**.** The Kaplan–Meier method gives an unbiased estimate of survival.

2. Definition of Terms

Log-extended Weibull survival regression model: This is a modified version of the Weibull distribution, which is a continuous probability distribution, it is a regression model which shows the relationship between the response variable and the explanatory variable, which is used in the analysis of survival rate of patient, studies or research about different disease and how long diagnosed patient can survive.

Kaplan-Meier: This statistical method which is used to present survival data, it gives a survival graph over time which shows the estimated percent survival of the subject(s) at each point in time. It is a method of summarizing survival data, which uses all the cases in a series not just those followed up until the selected cut-off. A survival probability is obtained When the surviving proportion is multiplied by the surviving proportions for each of the preceding periods, a probability is obtained which is called the survival probability**.**

Acute Myocardial Infraction; Acute myocardial infarction popularly known as Heart attack is a disease of the heart caused by blockage of the arteries supplying blood to the heart by fatty deposits.

3. MATERIALS METHODOLOGY

Extended Weibull distribution

Most generalized Weibull distributions have been proposed in reliability literature to provide a better fitting of certain data sets than the traditional two -or -three-parameter Weibull model. A very complicated generalized Weibull distribution often diminishes the probability of interpreting the parameters and a generalization that has more than three parameters is undesirable. The Xie et al.(2002) introduced a three-parameter Weibull distribution, the so-called the extended Weibull distribution, with cumulative distribution function (CDF) defined by ;

$$F(t) = 1 - \exp\left\{\lambda\alpha\left\{1 - \exp\left\{\frac{t}{\alpha}\right\}^{\beta}\right\}\right\} \tag{1}$$

Where λ > 0 and α > 0 are scale parameters and β > 0 is a shape parameter, the probability density function (PDF) is obtained by differentiating the cumulative distribution function (CDF) and we obtain the probability density function (PDF);

$$f(t;\lambda,\alpha,\beta) = \lambda\beta\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[\left(\frac{t}{\alpha}\right)^{\beta} + \lambda\alpha\left(1 - \exp\left\{\left(\frac{t}{\alpha}\right)^{\beta}\right\}\right)\right];\ t;\lambda,\alpha,\beta > 0 \tag{2}$$

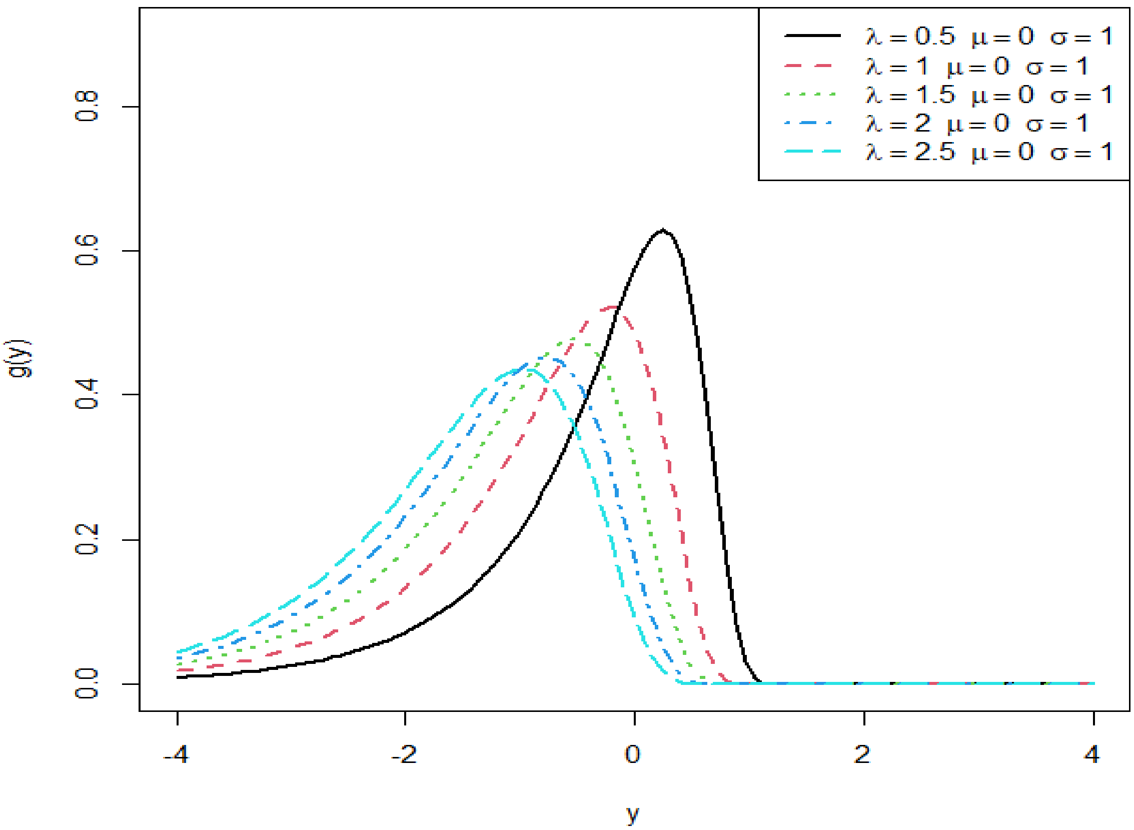

Fig.1: Cumulative Distribution Function plot of the Log extended Weibull distribution.

4. STATISTICAL PROPERTIES

The corresponding survival function reduces to;

$$S\left(y;\lambda,\sigma,\mu\right)=\exp\left\{\lambda\exp\left(\mu\right)\left[1-\exp\left[\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right]\right\} \tag{3}$$

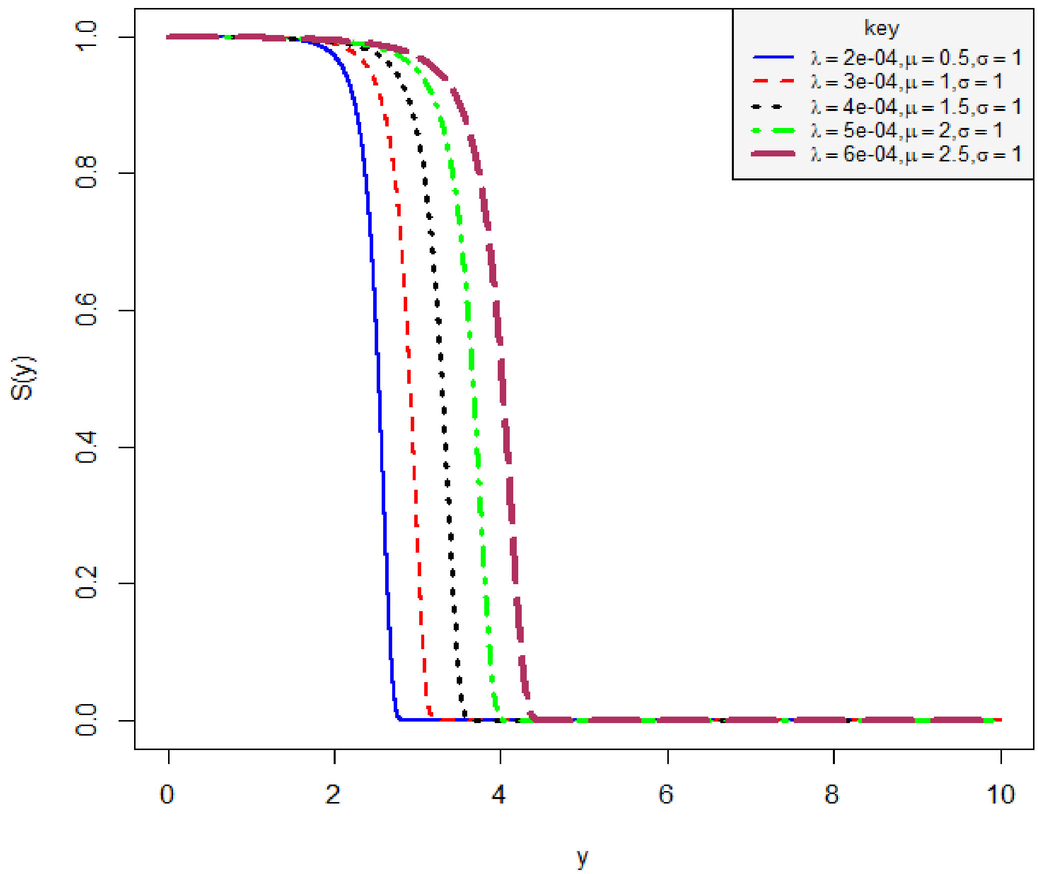

Fig.1: Survival Function Plot of the Log-Extended Weibull Distribution.

Kaplan-Meier survival probability

The Kaplan-Meier estimator of S(t) is given as ;

$$\hat{S}_n(t) = \prod_{\{i:T_{(i)} \le t\}} \left(1 - \hat{Q}_i\right) \tag{4}$$

$$= \prod_{\{i:T_{(i)} \le t\}} \left(\frac{n-i}{n-i+1}\right)^{\sigma_i} \tag{5}$$

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,

$$\frac{\hat{S}(t) - S(t)}{\sqrt{Var\left(\hat{S}(t)\right)}} \sim N(0,1) \tag{6}$$

Then for a given t, this would lead to asymptotic (1-α) per cent C.I for S(t):.

5. RESULTS AND DISCUSSION

The Acute Myocardial Infarction Survival Regression Model

Table 1: MLE of parameters of the Log Extended Weibull Survival Regression Model on the Acute Myocardial Infarction

Coefficients	Estimate	Standard Error	Z value	Pr(z)
Λ	5.8301482	0.24809551	23.5	<2.20e-16
σ	0.00064642	0.00013395	4.826	1.393e-07
β_0	5.0022816	0.16569982	30.189	<2.20e-16
β_1	-0.32749675	0.17227463	-1.901	0.0573
β_2	1.44687725	0.0206364	70.113	<2.20e-16
β_3	4.85133313	0.0272112	178.284	<2.20e-16
β_4	4.63566776	0.01309437	354.02	<2.20e-16
β_5	3.3929746	0.1787942	18.977	<2.20e-16
β_6	0.76901583	0.00112739	682.122	<2.20e-16
β_7	-2.63803271	0.00406726	-648.602	<2.20e-16
β_8	1.04847066	0.00234495	447.118	<2.20e-16
β_9	-4.29892724	0.17323942	-24.815	<2.20e-16

The Acute Myocardial Infarction (AMI) Dataset and its Exposure Variables

The acute myocardial infarction (AMI) survival data collected over 31 years from patients admitted to hospitals in Worcester, Massachusetts, represents a long-term, population-based study aimed at understanding trends in heart attack outcomes, survival rates, and treatment over time. The clinical trial related to the Worcester Heart Attack Study and its acute myocardial infarction (AMI) survival data is a longitudinal observational cohort study. The dataset contains $n = 100$ Myocardial infarction patients, where 49% of them are censored and 51% are uncensored. The response variable y is the natural logarithm of the observed survival time (in days); d = censoring (0 = alive at study end or lost to follow-up; 1 = death due to Acute myocardial infarction); with the following exposure variables: The response variable y is the natural logarithm of the observed survival time (in days); x1 = Patients age between 32 – 51; x2 = patients age between 52 – 71; x3 = patients age between 72 – 92 x4 = Male (1 is the indicator for male); x5 = Female (0 is the indicator for Female); x6 = Body mass index (BMI) of the patients greater than 18.5; x7 = Body mass index (BMI) of patients between 18.5 – 24.9; x8 = Body mass index (BMI) of the patients between 25 – 29.9; and x9 = Body mass index (BMI) of the patients at least $30kg/m^2$.

Table 2: Kaplan Meier Survival Probability Curve for Acute Myocardial Infarction Dataset.

Survival Time	Number at Risk	Number of Events	Survival Probability	Standard Error	95% Lower C.I	95% Upper C.I
6	100	2	0.98	0.014	0.9529	1
14	98	1	0.97	0.0171	0.9371	1
44	97	1	0.96	0.0196	0.9224	0.999
62	96	1	0.95	0.0218	0.9082	0.994
89	95	1	0.94	0.0237	0.8946	0.988
98	94	1	0.93	0.0255	0.8813	0.981
104	93	1	0.92	0.0271	0.8683	0.975
107	92	1	0.91	0.0286	0.8556	0.968
114	91	1	0.9	0.03	0.8431	0.961
123	90	1	0.89	0.0313	0.8307	0.953
128	89	1	0.88	0.0325	0.8186	0.946
148	88	1	0.87	0.0336	0.8065	0.938
182	87	1	0.86	0.0347	0.7946	0.931
187	86	1	0.85	0.0357	0.7828	0.923
189	85	1	0.84	0.0367	0.7711	0.915
274	84	2	0.82	0.0384	0.7481	0.899
302	82	1	0.81	0.0392	0.7366	0.891
363	81	1	0.8	0.04	0.7253	0.882
374	80	1	0.79	0.0407	0.7141	0.874
451	79	1	0.78	0.0414	0.7029	0.866
461	78	1	0.77	0.0421	0.6918	0.857
492	77	1	0.76	0.0427	0.6807	0.848
538	76	1	0.75	0.0433	0.6698	0.84
774	75	1	0.74	0.0439	0.6588	0.831
841	74	1	0.73	0.0444	0.648	0.822
936	73	1	0.72	0.0449	0.6372	0.814
1002	72	1	0.71	0.0454	0.6264	0.805
1011	71	1	0.7	0.0458	0.6157	0.796
1048	70	1	0.69	0.0462	0.6051	0.787
1054	69	1	0.68	0.0466	0.5945	0.778
1172	68	1	0.67	0.047	0.5839	0.769
1205	67	1	0.66	0.0474	0.5734	0.76
1278	66	1	0.65	0.0477	0.5629	0.751
1401	65	1	0.64	0.048	0.5525	0.741
1497	64	1	0.63	0.0483	0.5421	0.732
1557	63	1	0.62	0.0485	0.5318	0.723
1577	62	1	0.61	0.0488	0.5215	0.713
1624	61	1	0.6	0.049	0.5113	0.704
1669	60	1	0.59	0.0492	0.5011	0.695
1806	59	1	0.58	0.0494	0.4909	0.685

1874	52	1	0.569	0.0497	0.4794	0.675
1907	47	1	0.557	0.05	0.4668	0.664
2012	34	1	0.54	0.0512	0.4488	0.651
2031	32	1	0.523	0.0523	0.4304	0.637
2065	28	1	0.505	0.0537	0.4098	0.622
2201	14	1	0.469	0.0607	0.3636	0.604
2421	13	1	0.433	0.0659	0.321	0.583
2624	6	1	0.361	0.0857	0.2262	0.575
2710	2	1	0.18	0.1345	0.0418	0.778

Discussion of the Kaplan-Meier Survival Probability

Table 2 presents a Kaplan-Meier survival probability curve for patients with Acute Myocardial Infarction (AMI). The Kaplan-Meier estimator is a non-parametric statistic used to estimate the survival function from lifetime data, particularly when some patients may be lost to follow-up or censored. Here's a sound interpretation of the table components:

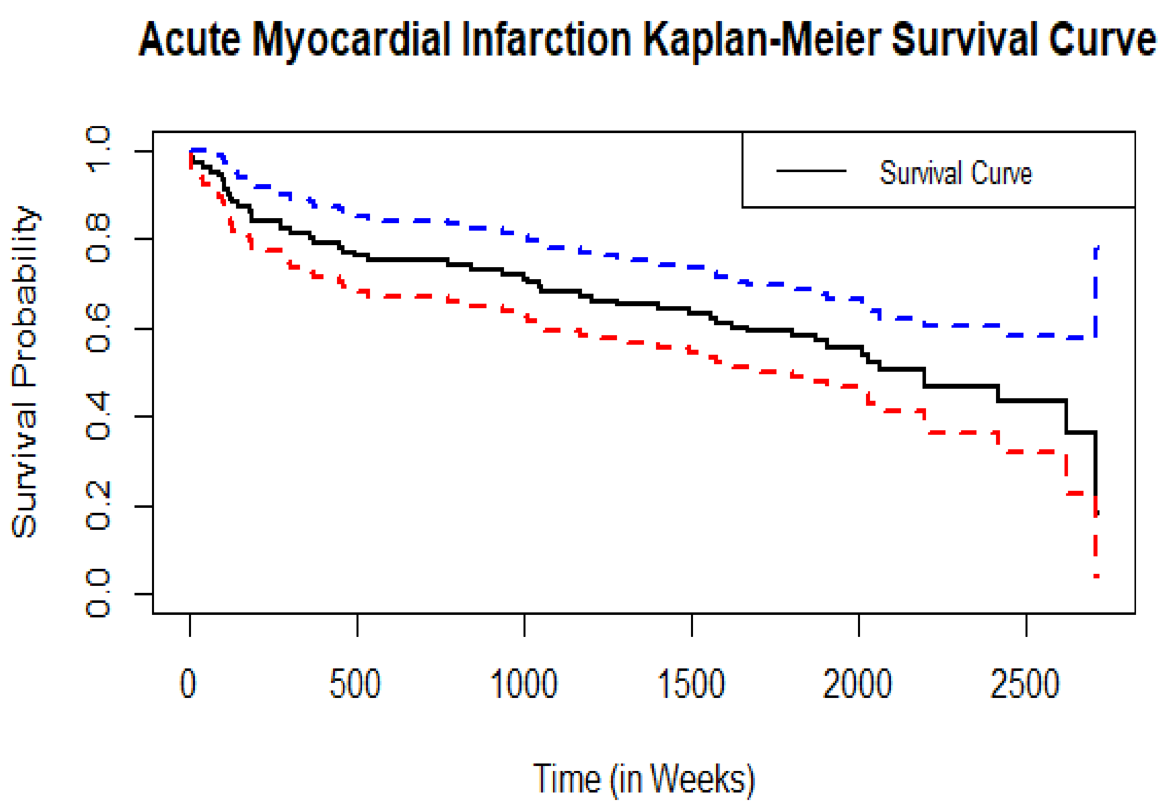

Fig. 2: plot of the Kaplan-Meier Survival curve

Interpreting the Curve:

A Kaplan-Meier curve is a visual representation of the probability of surviving over time in a group of individuals. In this case, Figure 4.1 shows the likelihood of survival for patients with acute myocardial infarction (heart attack). The **X-axis** represents Time (in Days); the **Y-axis** represents the survival probability (ranges from 0 to 1); and the **Survival curve** shows a step-like line that shows the proportion of patients who are alive at a given time together with its respective 95% upper and lower confidence interval.

Initial Survival: The curve starts at 1, indicating that all patients were alive at the beginning of the study.

Decreasing Survival: As time progresses (moving right on the x-axis), the curve typically declines, reflecting the loss of patients due to death or other factors.

Step-like Shape: The curve often has a step-like shape because it decreases at specific points in time when a patient dies. These steps represent a decrease in the survival probability.

6. CONCLUSION

The analysis of the Acute Myocardial Infarction (AMI) dataset using both the Log-Extended Weibull Survival Regression Model and the Kaplan-Meier Survival Estimator reveals critical insights into the survival patterns of AMI patients. The Log-Extended Weibull model identifies key covariates significantly influencing patient survival, highlighting specific risk factors with strong associations to mortality outcomes. In particular, parameters such as $\beta_2, \beta_3, \beta_4, \beta_6, and\ \beta_7$ these show highly significant effects on survival, suggesting the importance of these variables in patient prognosis.

The Kaplan-Meier survival curve shows that early survival after AMI is relatively high, with 98% of patients surviving beyond 6 weeks. However, survival steadily declines over time, with only 18% of patients surviving by the end of the study period at 2710 weeks (approximately 7.4 years). The consistent decrease in survival probability indicates the long-term risks associated with AMI, and the wide confidence intervals in later time points reflect growing uncertainty as fewer patients remain at risk.

REFERENCES

Biau, D. J., Latouche, A., & Porcher, R. (2007). Competing events influence estimated survival probability: when is Kaplan-Meier analysis appropriate?. *Clinical Orthopaedics and Related Research®, 462*, 229-233.

Bland, J. M., & Altman, D. G. (1998). Survival probabilities (the Kaplan-Meier method). *Bmj, 317*(7172), 1572-1580.

Cordeiro, G. M., Tahir, M. H., Vasconcelos, J. C. S., Ortega, E. M., & Hussain, M. A. (2021). A new extended log-Weibull regression: Simulations and applications. *Hacettepe Journal of Mathematics and Statistics, 50*(3), 855-871.

Etikan, I., Abubakar, S., & Alkassim, R. (2017). The Kaplan-Meier estimate in survival analysis. *Biom Biostat Int J, 5*(2), 00128.

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International journal of Ayurveda research*, *1*(4), 274.

Jager, K. J., Van Dijk, P. C., Zoccali, C., & Dekker, F. W. (2008). The analysis of survival data: the Kaplan–Meier method. *Kidney international*, *74*(5), 560-565.

Lacny, S., Wilson, T., Clement, F., Roberts, D. J., Faris, P. D., Ghali, W. A., & Marshall, D. A. (2015). Kaplan-Meier survival analysis overestimates the risk of revision arthroplasty: a meta-analysis. *Clinical Orthopaedics and Related Research®*, *473*, 3431-3442.

Mechanic, O. J., Gavin, M., & Grossman, S. A. (2017). Acute myocardial infarction.

Pollard, T. J. (2000). The acute myocardial infarction. *Primary Care: Clinics in Office Practice*, *27*(3), 631-649.

Silva, G. O., Ortega, E. M., & Cordeiro, G. M. (2009). A log-extended Weibull regression model. *Computational statistics & data analysis*, *53*(12), 4482-4489.