KAMPALA INTERNATIONAL UNIVERSITY
EXPLORING THE HEIGHTS

KIU
Journal of Science, Engineering and Technology

KJSET Research Article

Survival modelling of tuberculosis treatment: Insights from Kaplan-Meier and Cox proportional hazard models

Efuwape, Biodun T[1]., Badmus Nofiu I[1*]., Ogundeji, Rotimi K[2] and Amaury, de Souza[3]

[1]*Department of Statistics, Olabisi Onabanjo University, Ago-Iwoye, Nigeria, efuwape.biodun@oouagoiwoye.edu.ng*
[1*&2]*Department of Statistics, University of Lagos, Akoka, Yaba, Nigeria, E-mail: nibadmus@unilag.du.ng, rogundeji@unilag.edu.ng*
[3]*Universidade Federal do Mato Grosso do Sul, MS, Brazil, Amaury.souza@ufms.br*

Corresponding Author: nibadmus@unilag.du.ng

Paper history:
Received 29 March 2025
Accepted in revised form
07 May 2025

Keywords
Cox Proportional Model;
Isoniazid;
Kaplan Meier;
Pie Plot;
Rifampicin;
Tuberculosis.

Abstract

The study explores the crucial role of statistics in data science and analytics, particularly in statistical concepts and methods from the basis of data analysis, machine learning, and decision-making. The importance of statistical thinking in extracting insight from data and making informed decisions is also highlighted. Real-life survival data consisting of 203 tuberculosis patients with six variables, including age, gender, marital status, treatment (drug), censored (outcome), and time in days, is used to validate the study. Different analyses are carried out using R and Python software to depict summary statistics, plotting raw data, histograms, and percentages in pie plots, survival times for four different treatments include: Rifampicin, Isoniazid, Levofloxacin, and Bedaquiline, CoxPH model is fitted, and the Kaplan Meier survival curve for the treatment. The results reveal that Bedaquiline is given to 3 patients with no right-censored patient recorded, Levofloxacin was administered to 4 patients, only 1 is right censored, Isoniazid is administered to 36 patients, 8 are right censored, and Rifampicin was applied to 160 patients, and 8 are right censored. However, Rifampicin exhibits the highest survival rate, followed by Isoniazid, while Levofloxacin and Bedaquiline indicate greater uncertainty through their curves and shaded regions in the survival curve.

Nomenclature and units

$Coef$	Coefficient
$CoxPH$	Cox Proportional Hazard
$Exp(Coef)$	Exponential Coefficient
$f(t)$	Probability Density Function
$F(t)$	Distribution Function
$H(t)$	Hazard Function
$S(t)$	Survival Function
TB	Tuberculosis
T	Survival Time

1.0 Introduction

In today's data-driven world, Statistics play a central role in data science and quantitative analysis, providing essential tools for modeling, inference, and interpretation of complex phenomena. This study reinforces the applicability of statistics by exploring a clinical survival dataset of 203 tuberculosis patients, aiming to identify key factors influencing the effectiveness of four different treatments: Rifampicin, Isoniazid, Levofloxacin, and Bedaquiline. Additionally, tuberculosis infection and disease are treated with standard and special antibiotic drugs listed above. These drugs have to be taken daily for two to six months non-stop without medical advice. (Li et al., 2017, Mase and Chorba, 2019, (Bakare et al., 2022, and (Al-Karawi, Kadhim and Kadum, 2023). Several research studies are ongoing to combat tuberculosis globally, and these have saved about 79 million lives since 2000. (WHO, 2025). Recently, tuberculosis has become one of the world's leading causes of death after COVID-19, with a total of 1.25 million people dying from tuberculosis in 2023. However, the epidemic is among the health focus of the United Nations Sustainable Development Goals (SDGs) to eradicate by 2030. (Duarte et al., 2021, Gill et al., 2022, Rikochi, Musa and Olowolafe, 2023 and WHO, 2025). The analysis is conducted using advanced statistical techniques, such as the Cox Proportional Hazards Model (CoxPH) and Kaplan-Meier survival curves, to compare treatment efficacy and provide insights that can aid clinical decision-making. The chosen methods were selected for their capacity to model time-to-event data (in this case, survival time) and to handle censored data, a common feature in survival studies. The Kaplan-Meier curve was employed to estimate the survival function over time, while the CoxPH model was used to assess the impact of covariates on proportional hazards. In literature, different authors have researched and modelled on tuberculosis disease data (clinical data) in various ways such as parametric model by (Daniel, Lasisi and Banister, 2020), modeling as survival data by (Collett, 2023), a Cox proportional hazards by (Rexy and Rayalu, 2024), exponential and Weibull ATF models by (Akor et al., 2025) among others. The results indicate that Rifampicin and Isoniazid exhibit the best survival rates, with Rifampicin standing out as the most effective treatment, given the relatively low number of censored patients compared to other treatments. In contrast, Levofloxacin and bedaquiline displayed greater uncertainty in the survival curves, suggesting the need for larger sample sizes for validation. Some tools in statistics, but useful in data science and analytics, are R, Python, SPSS, SAS, Excel, and so on.

1.1 Survival Analysis

Survival analysis is one of the most significant applications of statistics in clinical trials, medical, and health sciences. It is also essential for modeling time-to-event data in various fields, like healthcare, engineering, finance, and so on. Estimating parameters from survival data, such as the Weibull Gamma and other distributions, is utilized instead of the normal distribution, allowing for modeling hazard rates and survival probabilities and facilitating predictions and decision-making (Klein et al., 2013). When describing survival statistics, the terms survival function, hazard function, and cumulative hazard function are often employed. Most people know that the proportional hazard model examines the connection between survival and the variables (Cox, 1972). It is often used in clinical trial analyses. The baseline hazard is a constant quantity that is unaffected by the variables and corresponds to an intercept. Using maximum likelihood, inferences about the baseline risks and the effects of explanatory factors are drawn from lifetime data in survival analysis (Groeneboom, Jongbloed, and Witte, 2020).

1.2 Tuberculosis (TB)

Tuberculosis (TB) remains a global public health problem and one of the top ten leading causes of death worldwide, with developing countries bearing the highest burden (WHO, 2022). In 2018, Nigeria was listed as first in Africa and sixth among the 30 countries with the highest TB burden (WHO, 2022). Unfortunately, the problem of TB in Nigeria has been complicated by the emergence and spread of drug-resistant TB and a high burden of HIV/AIDS NTBLCP, 2017; WHO, 2018). Here, we fit real-life survival data on tuberculosis to establish our facts using both R and Python software as statistical tools. The motivations for this are as follows: it helps in extracting insight from the data, making informed decisions, identifying patterns and trends, and quantifying uncertainty.

The paper is divided as follows: Section two contains the material and methods. Section three contains the analysis of data and interpretation, while section four contains the concluding remarks.

2.0 Materials and Methods

2.1 Survival Function

Survival analysis depends on the survival function $S(t)$, and it is known as the complementary cumulative distribution function. This function is a property of any random variable that maps a set of events, usually associated with mortality or failure of some system, onto time. It captures the

probability that the system will survive beyond a specified time.

Let T represent survival time. T is regarded as a random variable with a cumulative distribution Function:

$$F(t) = P(T \leq t) \quad (1)$$

and probability (event) density functions $f(t)$, $f(t) = F'(t)$. Given the probability that the event has occurred by duration t:

$$f(t) = \lim_{\Delta t \to 0} \frac{p(t \leq T \leq t + \Delta t)}{\Delta t} \quad (2)$$

The survival function, denoted as $S(t)$, is defined as follows for a random variable T representing survival time:

$$S(t) = P(T > t) = 1 - F(t) \quad (3)$$

where, $f(t)$ and $F(t)$ in equations (1) and (2) above represent the probability density and cumulative density functions of the specific distribution under consideration. The expression in equation (3) is survival time that quantifies the probability of survival beyond time t. It is important to note that $S(0) = 1$, and as time t approaches infinity, $S(t)$ approaches 0. The survival function typically forms a decreasing curve and can be estimated using methods like the Kaplan-Meier method discussed in section 2.4 below. The function can be used for two basic reasons: (i) can determine a patient's probability of surviving to time *t,* and (ii) can determine the % that survive to time *t*.

2.2 Hazard Function

In a set comprising individuals susceptible to a specific event, denoted as $R(t)$ (the risk set), or individuals who have not yet encountered the event by time t, the probability of an individual within this risk set facing the event within a short time interval $(t, t + \Delta t)$ is represented as $h(t)\,\Delta t$ Consequently, the hazard rate is formally defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t) \quad (4)$$

where T is non-negative and represents the future lifetime of an individual. In contrast to the survival function, which consistently exhibits a decreasing trend for all forms of survival data, the hazard function can assume various non-negative shapes, and its specific shape depends on the characteristics of the given survival data. Alternatively, the hazard function can also be expressed about the cumulative hazard function, denoted as $H(t)$.

$$h(t) = \int_0^t h(u)du \quad (5)$$

The term "cumulative" is employed because the function represents the accumulation of hazard over time.

2.3 Relationship Between Hazard and Survival Function

The relationship between them is obtained from (4) and is given by

$$\mathrm{h}(\mathrm{t}) = \mathrm{H}_0(\mathrm{t}) = \frac{1}{S(t)} \lim_{\Delta \mathrm{t} \to 0} \frac{\mathrm{S}(\mathrm{t}) - \mathrm{S}(\mathrm{t} + \Delta \mathrm{t})}{\Delta \mathrm{t}} = -\frac{S'(t)}{S(t)}$$

$$H(t) = \frac{S'(t)}{S(t)} - ln[S(t)] \quad (6)$$

where $H(t) = \int_0^t h(u)du$ and $H'(t)$ denotes the first derivative of the cumulative hazard function.

2.4 Kaplan Meier Estimator

The Kaplan-Meier estimator, alternatively known as the product limit estimator, is introduced by Kaplan and Meier in 1958. It offers a straightforward and rapid way to estimate the survival function, even when censoring is involved. It relies on the precise failure times (Kaplan and Meier, 1958; Collet, 2013), (Lowry and Kaplan, 2016 and Hanagal, 2011).

2.4.1 The Cumulative Hazard Function

Taking from expression (6), if $S'(t)$ is the Kaplan-Meier estimate of the survival function, then the following equation is an estimate of the cumulative hazard function:

$$H(t) = \sum_{i=1}^{m} ln\left(1 - \frac{d_i}{n_i}\right) \quad (7)$$

where d_i is the number of events at time t_i, and n_i is the number of individuals at risk (not censored) just before time t_i.

Then, from the Taylor series expansion $n\left(1 - \frac{d_i}{n_i}\right) = \frac{d_i}{n_i} - \left[\frac{d_i}{n_i}\right]^2 + \cdots \approx -\frac{d_i}{n_i}$ by ignoring higher-order terms. The estimate of the cumulative hazard function is, therefore, given as:

$$H(t) = \sum_{i=1}^{m} \frac{d_i}{n_i} \quad (8)$$

2.5 The Cox-proportional hazard model

The Cox-proportional hazard model is a broader and more versatile approach to modeling hazard and survival functions, as it does not impose distributional assumptions on the baseline hazard. This model was developed by Cox in 1972. It is in the form:

$$h(t|X) = h_0(t)exp(X^T\beta) \quad (9)$$

The assessment of how the covariates influence survival time is quantified using the hazard ratio, denoted as HR. Let's consider a categorical variable with two levels: $X = 1$ and $X = 0$. In this context, the hazard ratio for these two groups is articulated as:

$$HR = \frac{h(t|X=1)}{h(t|X=0)} = \exp(\beta) \quad (10)$$

An HR value of 1 indicates that individuals in both categories face an equal risk of experiencing the event. Conversely, when HR is greater than 1, it suggests that individuals in the first category ($X = 1$) have a higher risk of experiencing the event, while an HR less than 1 indicates that individuals in the second category ($X = 0$) face a heightened risk of experiencing the event.

The Cox-proportional hazard model relies on the assumption of proportional hazards. Consequently, the model may not be suitable for situations where this assumption is violated.

3.0 Results and Discussions

The data used contains 203 tuberculosis patients with six variables, including age, gender (status), marital Status, treatment (4 drugs such as Bedaquiline, Levofloxacin, Isoniazid, and Rifampicin), outcome (censored), and time (in days). The drugs were administered to different numbers of patients in the study. Bedaquiline was given to 3, Levofloxacin to 4, Isoniazid to 36, and Rifampicin to 160 patients. Therefore, data visualise like scatter plots, histograms, pie plots, Kaplan-Meier survival curves, tables, and their results were done by R and Python software. (R Code, 2022), (R Code Team, 2023) and Python, 2022). These are shown in Figure 1 below:

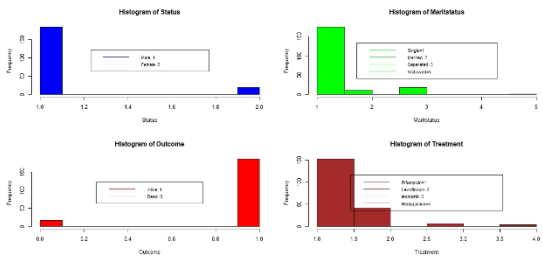

Figure 1 Histogram Plots for Status, Marital Status, Outcome, and Treatment

Figure 1 depicts the histogram of variables with number of patients: status/gender (male = 184 and female = 19), marital-status (single = 174, married = 10, seperated = 18, divorce = none, and widowed = 1), outcome/censored (alive = 186 and dead = 17), and treatment (Rifampicin = 153, Isoniazid = 21, Levofloxacin = 4 and Bedaquinine = 3).

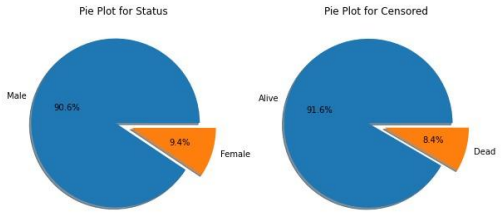

Figure 2 The Pie Plot for Status (Gender) and Censored

Figure 2 above shows the pie plot for gender and censored with their percentage. Gender comprises: male with 90.6% and female has 9.4%, and censored consists: alive with 91.6% and dead with 8.4%.

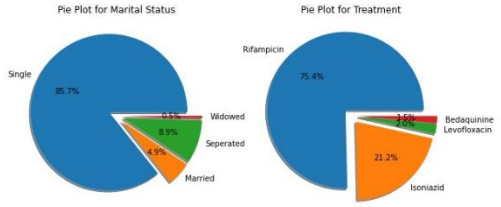

Figure 3 The Pie Plot for Marital Status and Treatment

Also, Figure 3 above exhibits the pie plot for marital status and treatment with their percentage. marital-status constitutes: single (85.7%), married (4.9%), separated (8.9%), and widowed (0.5%), and treatment makes up: Rifampicin (75.4%), Isoniazid (21.2%), Levofloxacin (2.0\% %), and Bedaquiline (1.5%).

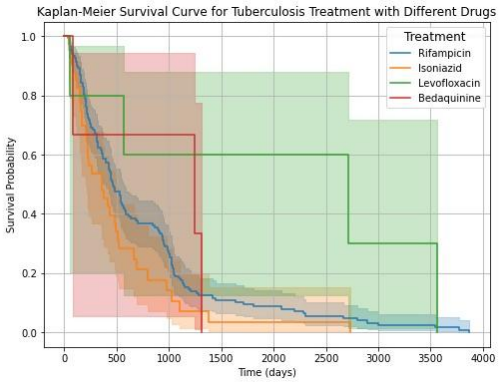

Figure 4 The Kaplan-Meier Survival Curve for Tuberculosis Treatment

The x-axis represents time, and the y-axis shows the survival probability (ranging from 0 to 1). The colored line represents a group of individuals being studied. The survival curve declines as events (such as deaths or failures) occur within the group. Group 1 (blue) has the highest survival rate initially, but as time progresses, the survival probability declines. The shaded regions around the survival curves represent the confidence intervals (often 95%) for the survival estimates. The wider shaded regions indicate greater uncertainty in the estimate due to a smaller sample size or more variability in survival times. For instance, the shaded area for groups 3 and 4 (orange and red) is wide at later times, which means high uncertainty. The "steps" in the curves occur when an event (such as death or failure) happens. When no events happen during a period, the line remains flat. A sharp drop in the curve indicates multiple events happening in close succession. Group 3 (orange colour) shows a sharp drop around the middle of the plot, and this indicates a significant number of events happening within a short time frame.

Table 1. Descriptive Summary Statistics of Tuberculosis Dataset

Statistics	Age	Gender	Marital Status	Treatment	Outcome	Time
Min	23.00	1.00	1.00	1.00	0.00	34.00
Mean	49.30	1.09	1.25	1.31	0.92	679.61
Std	12.93	0.29	0.65	0.60	0.28	798.52
25%	39.00	1.00	1.00	1.00	1.00	199.00
50%	48.00	1.00	1.00	1.00	1.00	429.00
75%	59.00	1.00	1.00	1.00	1.00	954.00
Max	83.00	2.00	5.00	4.00	1.00	3871.00

Table 1 contains summary statistics for each variable considered in the study. The statistics are: minimum, mean, standard error, 1st quartile, median, 3rd quartile, and maximum.

Table 2. Kaplan-Meier Survival Analysis for Bedaquiline Treatment

Time (days)	No of Risk	No of Event	Survival Probability	Standard Error	95% CI Lower	95% CI Upper
80	3.000	1.000	0.667	0.272	0.300	1.000
1241	2.000	1.000	0.333	0.272	0.067	1.000
1313	1.000	1.000	0.000	NaN	NA	NA

In Table 2 above, the survival analysis of Bedaquiline treatment for 3 patients has no death of patient recorded during the event, showing that: at 80 days, 66.7% of the patients survived at the initial survival rate, and the survival probability is 0.667. At 1241 days, survival probability decreased to 0.333 (33.3% patients) decline in survival probability. While at 1313 days, all patients experienced the event (death), resulting in a survival probability of 0.000, during the long-term outcome. The standard error and confidence intervals were NaN at 1313 days, likely due to the absence of remaining subjects. Moreover, Bedaquiline treatment depicts moderate initial effectiveness, wide confidence intervals show significant uncertainty due to the very small sample size (3 patients), and survival probability decreases over time.

Table 3. Kaplan-Meier Survival Analysis for Levofloxacin Treatment

Time (days)	No of Risk	No of Event	Survival Probability	Standard Error	95% CI Lower	95% CI Upper
51	5.000	1.000	0.800	0.179	0.516	1.000
570	4.000	1.000	0.600	0.219	0.293	1.000
2713	2.000	1.000	0.300	0.239	0.063	1.000
3564	1.000	1.000	0.000	NaN	NA	NA

Table 3. The survival analysis of Levofloxacin (drug) for the treatment of 4 patients, of whom 1 died. This shows that, at 51 days, 80% of the patients survived (survival probability of 0.800) at the initial survival rate. Gradual decline occurred when survival probability decreased over time, with 60% surviving at 570 days and 30% at 2713 days. Meanwhile, at 3564 days, all patients experienced the event (death), resulting in a survival probability of 0.000 as a long-term outcome. The standard error and confidence intervals were NaN, likely due to the absence of remaining subjects. In addition, Levofloxacin treatment shows limited effectiveness, with 80% survival at 51 days. Also, wide confidence intervals indicate uncertainty due to the small

sample size, and survival probability decreases over time, suggesting potential treatment limitations.

Table 4. Kaplan-Meier Survival Analysis for Isoniazid Treatment

Time (days)	No of Risk	No of Event	Survival Probability	Standard Error	95% CI Lower	95% CI Upper
46	42.000	1.000	0.976	0.024	0.931	1.000
49	41.000	1.000	0.952	0.033	0.990	1.000
50	39.000	1.000	0.928	0.040	0.853	1.000
70	38.000	1.000	0.904	0.046	0.818	0.998
“	“	“	“	“	“	“
“	“	“	“	“	“	“
“	“	“	“	“	“	“
1703	3.000	1.000	0.071	0.047	0.019	0.261
1274	2.000	1.000	0.035	0.034	0.005	0.237
2734	1.000	1.000	0.000	NaN	NA	NA

Table 4 consists of survival analysis of Isoniazid (drug) treatment of 36 patients, and 8 are right censored, revealing the initial survival probability at 46 days. Approximately 97.6% of patients were estimated to have survived (survival probability = 0.976). By 70 days, the survival probability decreased to 0.904, of patients were estimated to have survived as the survival probability declined. Also, by 2734 days, all patients had experienced the event (death), resulting in a survival probability of 0.000 at event occurrence. The standard error and confidence intervals were NaN, likely due to the absence of remaining subjects. It is discovered that as the days increased, the number of risks, survival probability, and 95% lower and upper bounds decreased. Summarily, Isoniazid treatment is effective in the short term, with high survival probabilities at early stages. Nevertheless, the survival probability decreases over time, indicating that the treatment’s effectiveness may decrease and cause the disease to progress.

Table 5. Kaplan-Meier Survival Analysis for Rifampicin Treatment

Time (days)	No of Risk	No of Event	Survival Probability	Standard Error	95% CI Lower	95% CI Upper
34	152.000	1.000	0.993	0.007	0.981	1.000
43	151.000	1.000	0.987	0.009	0.969	1.000
47	150.000	2.000	0.974	0.013	0.949	0.999
57	148.000	1.000	0.967	0.014	0.939	0.996
“	“	“	“	“	“	“
“	“	“	“	“	“	“
“	“	“	“	“	“	“
3548	3.000	1.000	0.016	0.011	0.004	0.062
3773	2.000	1.000	0.008	0.008	0.001	0.055
3871	1.000	1.000	0.000	NaN	NA	NA

Table 5. The survival analysis of Rifampicin (drug) treatment for 36 patients, with 8 right censored. At high initial survival probability, by 34 days, approximately 99.3% of patients were estimated to have survived (survival probability = 0.993). There is are gradual decline in survival probability by 57 days, the survival probability decreased to 0.967, indicating that about 96.7% of patients were estimated to have survived. However, at 3871 days, all patients had experienced the event (death), resulting in a survival probability of 0.000 at event occurrence. The standard error and confidence intervals were NaN, likely due to the absence of remaining subjects. In addition, it also reveals that as days increase, the number of risks, survival probability, and 95% lower and upper bounds are decreasing. In summary, Rifampicin treatment shows high efficacy in the short term, with a survival probability of 99.3% at 34 days. The survival probability gradually decreases over time, indicating that the treatment’s effectiveness may decrease or the disease may continue. Hence, the narrow confidence intervals at early stages, e.g., 34 and 57 days, suggest relatively precise estimates of survival probability.

Table 6. Cox Proportional Hazard Model for Treatment and Outcome

Covariate	Age	Gender	Marital Status	Time
Coef	0.004	-2.420	0.042	-0.000
Exp(Coef)	1.004	0.089	1.043	0.999
Se(Coef)	0.006	0.594	0.109	0.000
Coef 95% Lower	-0.007	-3.582	-0.171	-0.000
Exp (Coef 95% Lower)	0.014	-1.255	0.256	0.000
Coef 95% Upper	0.993	0.003	0.843	0.999
Exp (Coef 95% Upper)	1.014	0.285	1.292	1.000
Cmp	0.000	0.000	0.000	0.000
Z	0.636	-4.075	0.389	-0.053
p-value	0.525	0.000	0.697	0.958
-log(p)	0.930	14.405	0.521	0.066

The results in Table 6 are from the Cox proportional hazards model for treatment outcomes of the 203 tuberculosis patients. The model is used to estimate the effect of several covariates on the hazard or risk of an event happening (e.g., death, failure) and so on. The breakdown of the analysis goes thus:

A coefficient of 0.004 shows a small positive coefficient, indicating that age slightly increases the hazard. The exp(coef) reveals that the hazard ratio is 1.004, meaning that for every one-year increase in age, the hazard increases by about 0.40%. Then, the Confidence Interval CI for the hazard ratio is (0.993, 1.014), which includes 1, indicating no significant effect. The p-value (0.525) is quite large, indicating that Age is not a significant predictor of the hazard.

As for gender, a coefficient of (-2.420) has a large negative coefficient, which indicates that a higher Status value is associated with a decreased hazard. Exp(coef) is the hazard ratio of 0.089, which is significantly less than 1, and implies that a higher Status is associated with a 91% reduction in the hazard. The CI for the hazard ratio is (0.003, 0.285), which does not include 1, suggesting a statistically significant effect. Then, the p-value is highly significant (less than 0.05), meaning that gender is a statistically significant predictor of the hazard and has a strong effect.

Furthermore, the coefficient of marital status is 0.042, a small positive coefficient that suggests that being in a certain marital status slightly increases the hazard. Exp(coef) yields 1.043, which indicates that the hazard ratio is slightly above 1.000; this indicates that marital status increases the hazard by 4.3%. The CI for the hazard ratio is (0.843, 1.292), which includes 1, indicating no significant effect. The p-value is (0.697) very high, indicating that marital status is not statistically significant.

Meanwhile, time discloses the following results as its coefficient gives -0.000, indicating that the coefficient is almost zero and that Time has virtually no effect on the hazard. The exp(coef) is 0.999, the hazard ratio is almost 1, indicating no real change in hazard with time. The CI for the hazard ratio is (0.999, 1.000), which includes 1.000, confirming no significant effect. The p-value = 0.958 is extremely high, and this suggests that Time is not statistically significant.

3.1 Summary of Kaplan-Meier Plot

Group 1 (blue colour) and group 2 (green colour) have relatively similar survival curves for most of the timeline, though group 1 shows slightly better survival in the early stages. Group 3 (orange colour) and group 4 (red colour) both in Figure 4, show different patterns, also group 3 experienced a sharp drop, while group 4 had the most variability (as indicated by the wide confidence intervals in Table 2). Group 4 has the steepest drops in survival probability early on. Groups 1 and 2 appear to have better survival outcomes, as indicated by their higher survival probabilities at various times. Groups 3 and 4 have worse survival outcomes, with sharp drops in survival probability and larger uncertainty (as seen in the wide confidence intervals in Tables 2 and 3). In summary, Bedaquiline fitted 3 observations with zero right-censored observations, Levofloxacin fitted 4 observations with 1 right-censored observation, Isoniazid fitted 36 observations with 8 right-censored observations, and Rifampicin fitted 160 observations with 8 right-censored observations.

3.2 Summary of Findings of Cox Proportional Hazard Model

The only statistically significant covariate in this analysis is Status (p-value = 0.000046), which shows a strong protective effect with a hazard ratio = 0.089, given a 91% decrease in hazard. Age, marital status, and Time do not have statistically significant effects on the hazard based on the provided results. Their p-values are greater than 5%, and their confidence intervals for the hazard ratios include 1, indicating no strong effect on survival. Thus, we conclude that Status is the key factor influencing survival in this model, while the other covariates do not show significant associations with the outcome. CoxPH fitted with 203 total observations, and 17 are right-censored observations. Statistics are crucial in the world of data. Without data, data science and analytics are rendered ineffective.

From the analysis, Rifampicin and Isoniazid emerged as promising treatments, primarily due to their stable survival curves and lower censorship rates. Visual inspection of the Kaplan-Meier curves and comparison of confidence intervals indicate that these treatments provide better survival outcomes, especially in the early stages of treatment. Notably, the group treated with Rifampicin showed less variability in survival curves, indicating consistent results and robust therapeutic efficacy.

Conversely, the uncertainty observed in the survival curves for Levofloxacin and bedaquiline may be related to the small sample size and the absence of censored observations for bedaquiline, which suggests a selection bias. Including a larger number of patients and evaluating these treatments in a multicenter study could provide more clarity on the actual effectiveness of these drugs in tuberculosis treatment. Moreover, other factors, such as bacterial resistance and drug co-administration, could be considered in future analyses to assess how these elements modulate treatment response.

In conclusion, this study demonstrates the practical applicability of statistical techniques in clinical data analysis, using the CoxPH model and Kaplan-Meier curves to evaluate the efficacy of different treatments for tuberculosis. Rifampicin and Isoniazid stood out as the best treatment options, while Levofloxacin and Bedaquiline require further investigation. The analysis provides valuable insights for clinical practice and highlights the importance of statistics as a cornerstone for data science and evidence-based decision-making. Future studies should address and discuss limitations. Also, considering the

use of more advanced statistical methods to obtain more precise and generalizable estimates, contributing to a better understanding of the factors that influence the survival of tuberculosis patients. Summarily, the Cox proportional hazard model identified e.g., age, gender, marital Status and time as significant predictors of mortality.

4.0 Conclusions

This study investigated the survival outcomes and predictors of mortality among patients treated with four different anti-tuberculosis medications: Bedaquiline, Levofloxacin, Isoniazid, and Rifampicin. The Kaplan-Meier plot in Figure 4 revealed varying survival probabilities for patients treated with these medications. Hence, the Cox proportional hazard model output stated in Table 6 identified significant predictors of mortality, providing insights into the factors that influence treatment outcomes.

Acknowledgements

The authors would like to express their gratitude to those who supported the success of the research.

Declaration of conflict of interest

The authors declare that there is no conflict of interest

References

Akor, A., Sadiq, I. A., Usman, A., Doguwa, S. I and Akor, L. O. (2025). Advanced Survival Modeling of Tuberculosis Patients: Insight from Exponential and Weibull AFT Models. *FUDMA Journal of Sciences, 9*(3), 169 – 182. doi:10.33003/fjs-2025-090

Al-Karawi, S. A., Kadhim, A. A and Kadum, M. M. (2023). Recent Advances in tuberculosis: A Comprehensive Review of emerging Trends in Pathogenesis, Diagnostics, Treatments, and Prevention. *Int. J. of Clin. Biochem Res, 10*(4), 262 – 269.

Bakare, A. A., Moses, V. Y., Beckely, C. T., Oluyemi, T. I., Ogunfeitimi, G. O and Adelaja A. A. (n.d.). The first-line Anti-tuberculosis drugs, and their fixed-dose combination induced abnormal sperm morphology and histological lesions in the testicular Cells of Male Mice. *Frnt Cell Dev. Biol, 10*, 1023413.

Collet, D. (2013). *Modelling Survival Data in Medical Research* (4th Edition ed.). CRC Press.

Cox, D. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B, 34*, 187-220.

Daniel, S., Lasisi, K. E and Banister, J. (2020). Application of Survival Analysis of TB Patients Using Parametric Model: A case Study of General Hospital Bayara. *Asian Journal of Probability and Statistics, 6*(4), 54 - 56. Retrieved from https://doi.org/10.9734/ajpas/2020/06i430189

Duarte, R., Aguiar, A., Pinto, M., Furtado, I., Tiberi, S and Lönnroth, K. (2021). Different disease, same challenges: Social determinants of tuberculosis and COVID-19. *Pulmonology, 27*(4), 38 - 44.

Foundation, P. S. (2022). Python. Retrieved from https://www.python.org

Gill, C. M., Dolan, L., Piggott, L. M and Mclaughlin, A. M. (2022). New developments in tuberculosis diagnosis and treatment. *Breathe (Sheff), 18*(1), 210149.

Groeneboom P, Jongbloed G, Witte B. I. (2020). Analysis of type I and II error rates of Bayesian and frequentist parametric and nonparametric two-sample hypothesis tests under preliminary assessment of normality. *Computational Statistics*, 101-109.

Hanagal, D. D. (2011). Modeling Survival Data Using Frailty Models. *Journal of Industrial and Applied Mathematics*. doi:10.1201/b10510

Kaplan, E. L and Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistics Association, 53*, 457-481. Retrieved from http://www.jstor.org/stable/2281868

Klein, J. P, Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd Edition ed.). United Kingdom: Springer. doi:10.1007/b97377

Li, J., Chung, P. H., Leung, C. L. K., Nishikiori, N., Chan, E and Yeoh, E. K. (2017). The strategic framework of tuberculosis control and prevention in the elderly: a scoping review towards End TB targets. *Infect Dis Poverty, 6*(1), 70.

Lowry, R., Kaplan-Meier, E. L. (2017). Survival Probability Estimates VassarStats. *Journal of the Practice of Cardiovascular Sciences*. Retrieved from Website for Statistical Computation, http://www.vassarstats.net/survival.html

Mase, S. R and Chorba, T. (2019). Treatment of Drug-Resistant Tuberculosis. *Clin. Chest Med, 40*(4), 75 - 95.

Rexy, D. C and Rayalu, G. M. (2024). Exploring Time-Dependent Factors in the Survival of tuberculosis Patients: A Cox Proportional Hazards Analysis. *Machine Intelligence Research, 18*(1), 556 – 565. Retrieved from https://machineintelligenceresearchs.com/index.php/mir/article/view/50

Rikochi, C. L., Musa, A. Z and Olowolafe, T. A. (2023). Implementation of WHO Guideline on Tuberculosis Infection, Prevention and Control in Kaduna State, Nigeria. *PAMJ-One Health, 12*(1). doi:10.11604/pamj-oh.2023.12.1.41035

Team, R. C. (2022). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna: R. Retrieved from https://www.R-project.org

Team, R. C. (2023). A Language and Environment for Statistical Computing: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

WHO. (2018). *Global Tuberculosis Report.* WHO. Retrieved 2025, from WHO/CDS/TB/2018.20

WHO. (2022). *Global Tuberculosis Report.* WHO. Retrieved from https://www.who.int/publications/i/9789240061729

WHO. (2025). *Tuberculosis.* Retrieved from https://www.who.int/news-room/fact-sheets/detail/tuberculosis