**A Weibull-Fréchet proportional hazard model with Applications to Tuberculosis Data**

**Abstract:**

*This paper extends the application of survival analysis by adopting a Weibull-Fréchet distribution within the framework of the Cox Proportional Hazard (PH) model. The study focuses on 319 Tuberculosis (TB) patients from National TB and Leprosy Center Hospital, Kaduna, Nigeria (NTLCHKN). It extends the Weibull-Fréchet distribution via it hazard function to a proportional Hazard model framework and estimated the parameters using maximum likelihood estimation with Newton Raphson optimization in R-software. In the analysis of NTLCHKN data, the Weibull-Fréchet Proportional Hazard (WFr) Model demonstrates superior fitting compared to the Weibull Proportional Hazard Model. Consequently, result from the model identifies age, gender, type of TB, smoking history, and comorbidities (hepatitis and HIV-AIDS) as significant predictors of TB mortality. This research contributes to the ongoing advancements in survival analysis, highlighting the efficacy of the distribution in modeling TB mortality and providing valuable insights for tailored interventions and improved patient outcomes in TB management. Thus, the model can be employed as preferred alternative choice for researchers and practitioners engaged in tuberculosis survival data or other phenomenon data on survival analysis.*

**Keywords:** Tuberculosis, Proportional Hazards, Cox’s model , Weibull-Fréchet distribution

**Introduction:**

Survival analysis plays a significant role in understanding time-to-event data, especially in medical, engineering, and social sciences. Proportional Hazards (PH) models are widely used in survival analysis, offering a flexible framework for modeling the hazard function. The Breslow-Cox model, also referred to as the Cox Proportional Hazard (PH) model or simply Proportional Hazard model, is widely recognized as a prominent approach for analyzing survival data. The non-parametric baseline hazard rate estimator of the Cox proportional hazard model is recognized for its significant adaptability in capturing temporal variations within the hazard rate, however this estimator is not smooth, and it yields an event time probability density that lacks correct normalization, rendering it unsuitable for making predictions (Barrett, 2014). The Cox proportional hazards model is widely used in epidemiological analyses of cohort data (Rothman, Greenland, & Lash, 2007). The Cox PH model is mathematically written as:

= (1)

where  is the baseline hazard function of distribution of interest and **β** is the vector of the regression parameters and **X** represent explanatory variables in the model. The baseline hazard refers to the hazard function applied to an individual when all covariates are set to zero. This means that it represents the inherent risk of experiencing the event of interest without considering any additional factors or characteristics of the individual. Since the baseline hazard is not assumed to be of a parametric form, Cox’s model is referred to as a semi-parametric model for the hazard function. As outlined in Cox's (1972) Proportional Hazards (PH) model, the assumption is that the risk within groups remains proportional over time. In cases where the hazard rate is not constant, the survival data suggests an assumption of hazard distribution across different groups (Cox, 1972).

Conversely, in the context of parametric PH models, the Weibull distribution stands out as the most commonly employed parametric time-to-event model (Diamoutene et al., 2021; Zhang, 2016). This distribution is very flexible and can be transformed to the exponential distribution. The shape and scale parameter for this distribution is denoted by and  respectively. For a Weibull distribution, suppose the survival time T ~ the hazard function of the Weibull distribution is given;

   (2)

The shape parameter affects the shape of the hazard function and hence can take different values. When the hazard function is increasing the shape parameter > 1 making it similar to l memory loss distribution. When the hazard function is decreasing the shape parameter < 1. The Weibull proportional hazard model considering the individual is given by

 (3)

Studies with an empirical focus, such as the study (Khan & Khosa, 2015) examine the formulation of proportional hazard (PH) models, which can be approached with or without assuming a probability distribution for survival times. They note that assuming a distribution leads to parametric models, while the semi-parametric Cox model is widely favored in survival analysis. However, parametric models may offer more efficient estimates under specific conditions. The authors highlight that only a few parametric models are closed under the PH assumption, with the Weibull being the most common, accommodating only monotone hazard functions. To address this limitation, they propose a generalized form of the log-logistic distribution within the PH family. This distribution shares properties with the log-logistic and approaches the Weibull in the limit, allowing it to handle both monotone and nonmonotone hazard functions. Through application to four datasets and a simulation study, they demonstrate the potential utility of this model in adequately describing various types of time-to-event data.

A study by (Faruk et al., 2018) investigated the determinants of the first birth interval (FBI) in Indonesia, considering it as a crucial indicator of women's fertility rates. They stated that the prolonged FBI signifies a lower fertility rate, as such a concern for populous countries like Indonesia aimed to curb overpopulation and to determine the significant factors influencing FBI and quantify their effects, the researchers employed survival analysis techniques. Specifically, they proposed the Weibull proportional hazards (PH) model, utilizing data from the 2012 Indonesian Demographic and Health Survey (IDHS), comprising 27,488 ever-married women aged 15-49. Comparing various approaches including Kaplan-Meier method, log-rank test, and PH model, the study revealed that the Weibull PH model outperformed others. Furthermore, the analysis demonstrated that factors such as age at first birth, place of residence, women's and husband's educational levels, contraceptive knowledge, wealth index, and employment status significantly influenced FBI data in Indonesia.

 Similarly, the study of (Hongxiang et al., 2020) emphasizes the important roles played by the Weibull distribution and parametric models, such as proportional hazards models, in survival data analysis. In their paper they introduced two-parameter, three-parameter, and four-parameter modified Weibull distributions, illustrating their applicability within proportional hazards models. Utilizing smoothed hazard estimates, the study suggests that Weibull and modified Weibull models are preferable for coronary heart disease (CHD) data analysis. The researchers fit Weibull and modified Weibull proportional hazards models to CHD data, concluding that the two-parameter modified Weibull model offers a superior fit compared to Weibull and other modified Weibull proportional hazards models. Hongxiang et al., (2020) also opined that the hazard function in the Weibull distribution fails to accommodate the complexities of certain hazard models and lacks a non-monotonic hazard rate, such as the bathtub shape.

**MATERIALS AND METHOD**

Numerous modified Weibull distributions have been introduced over the last three decades, aiming to overcome the limitations of the traditional Weibull distribution. This paper explores one of these modified Weibull distributions called the Weibull-Fréchet (WFr) distribution developed by (Afify et al., 2016). The Weibull distribution has the ability to model different types of reliability data exhibiting either skewed or symmetric distributional shape (Lawless, 2011). Meanwhile, the Frechet distribution, also known as the extreme value type II distribution, is commonly utilized to model extreme events with heavy tails (Balakrishnan & Aggarwala, 2000). The fusion of these distributions, known as the Weibull-Frechet (WFr) distribution, provides enhanced flexibility in capturing diverse survival patterns due to its additional location parameter (Afify et al., 2016). The WFr hazard function for some parameter value reveal that this function can be unimodal, decreasing or increasing, depending on the parameter values. The hazard function is given by

 (4)

The aforementioned introduced four-parameter Weibull-Fréchet distribution was initially proposed for simulation data and examined in real data without covariates. However, it has not been integrated into a parametric proportional hazards model and is rarely employed in the analysis of actual survival data. This paper suggests adopting the Weibull-Fréchet distribution for proportional hazards model. Additionally, a Tuberculosis (TB) case study of the National TB and Leprosy Center Hospital in Zaria, Kaduna State, Nigeria (NTLCHKN) data set will be employed, by considering sex, age, types of TB, smoking history, alcohol history, and comorbidity as TB risk factors. The study aims to compare the performance between the Weibull and Weibull-Fréchet distribution based on a proportional hazards model and investigate the risk factors associated with mortality of tuberculosis patients.

**PROPORTIONAL HAZARD REGRESSION**

First of all, suppose the survival time has a Weibull-Frechet distribution (WFr) where α is a scale parameter representing the characteristic life and β, a and b are shape parameters representing the different patterns of the WFr distribution, the baseline hazard function is given by

 (5)

The probability density function, survivor function and cumulative hazard function of the Weibull-Frechet distribution are, respectively,

 (6)

 (7)

 (8)

substituting (5) into equation (1) as the baseline hazard function at time *t* for the *i*th of n subjects is given by

 (9)

Equation (9) manifests that the survival time of the *i*th individual has a distribution. Therefore, the four-parameter Weibull-Frechet distribution has the proportional hazards property, thus can be refer to as Weibull-Frehect Cox proportional Hazard model (WFr CPH) or simply Weibull-Frechet proportional hazard model (WFr PH). Thus, the estimated survival function corresponding to the hazard function given in equation (9) can be derived as

 (10)

 (11)

 (12)

 (13)

And  represent the cumulative hazard function. Hence, the full likelihood of estimating parametric survival model can be written as either as a function of survival function and failure density function, or as a function of the survival function and hazard function. Thus, the likelihood of estimating our proposed WFr parametric survival model is formulated as a function combining the survival function and hazard function, that is

 (14)

 (15)

 (16)

where K is the set of all individuals where the event was observed (uncensored), and M is the set of all individuals where the event was not observed (censored). The corresponding loglikelihood, that will often be subsequently optimised, is as follows,

 (17)

Where  indicates the cumulative hazard up until time *ti.*. Thus, substituting equation (9) and (12) in equation (16) gives the log-likelihood estimation equation for the WFr Cox proportional Hazard model.

 (18)

 (19)

(20)

 (21)

 (22)

 (23)

 (24)

 (25)

According to (Collett, 2015) the log-likelihood estimates of the parameters in the PH model can then be found by maximizing this log-likelihood function using numerical methods. The optimal values of the parameters are given by numerically solving which can be carried out using a *flexsurv* package in R.

**RESULT AND DISCUSSION**

The study provides a thorough analysis of demographic and health-related variables associated with tuberculosis (TB) and its influencing factors. The age distribution emphasizes the prevalence of TB among individuals aged 35-55 (60%), indicating a critical focus for targeted interventions. The gender distribution reveals 64% males and 36% females in the study population, with a majority being single (82%). This underscores the necessity of understanding support systems, particularly for TB patients lacking familial support. Pulmonary TB cases constitute 84%, while extra-pulmonary cases make up 16%, highlighting diverse challenges in diagnosis and treatment. A notable proportion of participants exhibit a history of alcohol consumption (80%) and smoking (48%), emphasizing the need to address lifestyle factors impacting TB outcomes. Prevalent comorbidities, such as hepatitis (35%) and HIV-AIDS (36%), underscore the complex nature of TB management, necessitating comprehensive care strategies. In conclusion, the study enhances understanding of TB demographics, emphasizing the importance of gender-specific considerations, marital status, and lifestyle factors. The high prevalence of comorbidities, family TB history, and age distribution underscore the necessity for a holistic approach in TB treatment and control.

**Table 1: Patients Characteristics**

|  |  |  |  |
| --- | --- | --- | --- |
| **Covariates** | **Categories** | **Frequency** | **Percentage**  |
| Age | <35 years | 86 | 27% |
| 35-55 years | 191 | 60% |
| > 55 years | 42 | 13% |
| Gender | Female | 115 | 36% |
| Male | 204 | 64% |
| Type TB | Pulmonary | 269 | 84% |
| Extra Pulmonary | 50 | 16% |
| Alcohol history | No | 64 | 20% |
| Yes | 255 | 80% |
| Smoking history | No | 166 | 52% |
| Yes | 153 | 48% |
| Comorbidity | No comorbidity | 91 | 29% |
| Hepatitis | 113 | 35% |
| HIV-AIDS | 115 | 36% |

**MODEL COMPARISON**

The AIC and HQIC, serving as measures of model fit and complexity, collectively pointed towards the Weibull-Fréchet model as the optimal choice. This careful model selection process ensures that the chosen survival model not only aligns with the data but also strikes a balance between explanatory power and simplicity, enhancing the reliability of the subsequent survival analyses and interpretations in the context of the research at hand.

Fig 1 Comparison between Weibull-Fréchet PH and Weibull PH model



The Weibull-Fréchet PH model appears to track the Kaplan-Meier curve more closely especially over the later follow-up periods, suggesting a potentially better fit to the data than the standard Weibull PH model, suggesting better fit to the data. However, its lower Information Criterion values support its improved performance, indicating its superiority over the standard Weibull PH model.

**Table 2: Model comparison using AIC and HQIC**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model**  | **AIC** | **HQIC** | **Rank** |
| **Weibull Cox PH** | 870.6598 | 885.6966 | 2nd |
| **Weibull-Fréchet Cox PH** | 867.0686 | 885.1121 | 1st |

The Weibull and Weibull-Fréchet Proportional Hazard Model identifies age, gender, type of TB, smoking history, and comorbidities (hepatitis and HIV-AIDS) as significant predictors of TB mortality. These findings provide valuable insights for tailoring interventions and improving patient outcomes in TB management.

**Table 3: Multivariate Parametric Weibull and Weibull-Fréchet PH models for Tuberculosis data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Covariates** | **Categories** | **Weibull** | **Weibull-Fréchet** |
| **Haz. Ratio (P.value)** | **Haz.Ratio (P.value)** |
| Age Group | <35 years (Ref) |  |  |
| 35-55 years | 1.01 (0.0181) | 1.03 (0.0166) |
| > 55 years | 1.66 (0.0004) | 1.66 (0.0005) |
| Gender | Female (Ref) |  |  |
| Male | 1.05 (0.0108) | 1.05 (0.0107) |
| Type\_TB | Pulmonary |  |  |
| Extra Pulmonary | 1.27 (0.0001) | 1.25 (0.0001) |
| Alcohol history | No (Ref) |  |  |
| Yes | 0.0096 (0.3920) | 0.0675 (0.4240) |
| Smoking history | No (Ref) |  |  |
| Yes | 1.56 (0.0001) | 1.55 (0.0001) |
| Comorbidity | No comorbidity (Ref) |  |  |
| Hepatitis | 1.21 (0.0107) | 1.21 (0.0105) |
| HIV-AIDS | 1.37 (0.0029) | 1.38 (0.0027) |
|  |  |  |  |
|  | **Estimated Parameter of Weibull and WFr PH model** |
| Shape (a) |  | - | 0.000197 |
| Shape (b) |  | - | 0.138 |
| Shape (c) |  | 1.98 | 67.2 |
| Scale (α) |  | 0.00000406 | 13.9 |

The applications of the Weibull proportional hazards model in survival data analysis are common. This study offers an example of a scenario in which the Weibull-Fréchet proportional hazards model more accurately describes the TB survival data than Weibull PH model. In addition, to evaluate the adequacy of the fitted WFr Proportional Hazards model, a Cox-Snell residual analysis was conducted. The plot below displays the estimated cumulative hazard function of the Cox-Snell residuals against the residual values. Specifically, Figure 2 shows that the hazard function for the of the model successfully followed a straight line from the origin and are within the confidence interval of cumulative hazard values which reflected how well the model fit the data. This finding reveals that the WFr PH model fit the TB data well for all values of survival time.



**Figure 2: Weibull Fréchet Proportional Hazard Model Diagnosis**

**CONCLUSION**

In summary, this study provides a significant progress in survival analysis by incorporating the Weibull-Frechet (WFr) distribution as the baseline hazard function to Cox proportional hazard model, thereby extending the proportional hazard framework. Thus, the developed WFr PH model was applied to tuberculosis patient data from the National TB and Leprosy Center Hospital in Zaria, Kaduna Nigeria. When compared to the standard Weibull model, the Weibull-Fréchet Proportional Hazards model provided a better overall fit, enabling a deeper understanding of factors influencing TB mortality. Notably, variables such as age, gender, type of TB, smoking history, and comorbidities like hepatitis and HIV-AIDS were identified as significant predictors, emphasizing their important roles in patient prognosis. Clinically, these findings reinforce the need for targeted interventions focusing on these high-risk groups to improve survival outcomes. Additionally, the adaptability and enhanced precision of this modeling approach extend its utility far beyond tuberculosis research, providing a robust analytical framework for examining time-to-event data across diverse fields including medical research, epidemiological investigations, and engineering applications where understanding historical patterns of events is essential. In summary, this research contributes to both theoretical and practical dimensions by improving survival modeling methodologies and delivering evidence-based insights for enhanced tuberculosis management and comparable health conditions.

**Ethics declaration**

The study titled does not require ethical consent because it relies solely on secondary data that is fully anonymized. The data used in this research has been previously collected, de-identified, and made publicly available for academic and research purposes. Consequently, there is no risk of compromising the privacy or confidentiality of individual patients, and no direct interaction with human subjects is involved in this study.

**Disclaimer**

This paper is an extended version of a preprint document of the same author.

The preprint document is available in this link: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4879413>

[As per journal policy, preprint /repository article can be published as a journal article, provided it is not published in any other journal]

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.NIL

2.NIL

3.NIL

**Reference**

Afify, A. Z., Yousof, H. M., Cordeiro, G. M., Ortega, E. M. M., & Nofal, Z. M. (2016). The Weibull Fréchet distribution and its applications. *Journal of Applied Statistics*, *43*(14), 2608–2626. <https://doi.org/10.1080/02664763.2016.1142945>

Balakrishnan, N., & Aggarwala, R. (2000). A primer on statistical distributions. John Wiley & Sons.

Barrett, J. (2014). *Weibull-Cox proportional hazard model*. Mathematics, 36, 2803 - 2813.

Collett, D. (2014). Modelling Survival Data in Medical Research (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b18041>

Cox, D. R. 1972. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2): 187–202.

Diamoutene, A., Noureddine, F., Kamsu-Foguem, B., & Barro, D. (2021). Reliability Analysis with Proportional Hazard Model in Aeronautics. *International Journal of Aeronautical and Space Sciences*, *22*(5), 1222–1234. https://doi.org/10.1007/s42405-021-00371-1

Faruk, A., Setyo Cahyono, E., & Eliyati, N. (2018). Fitting The First Birth Interval in Indonesia Using Weibull Proportional Hazards Model. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, *5*(2), 55–63. https://doi.org/10.18860/ca.v5i2.4060

Hongxiang, L., Shan, F. P., & Baofeng, S. (2020). A comparative study of modified Weibull distributions in proportional hazards models. *AIP Conference Proceedings*, *2266*. <https://doi.org/10.1063/5.0018428>

Lawless, J. F. (2011). Statistical models and methods for lifetime data (Vol. 362). John Wiley & Sons.

Rothman, K. J., Greenland, S., & Lash, T. L. (2007). 3 Epidemiologic Study Designs. *Handbook of Statistics*, *27*, 64-108.

Khan, S. A., & Khosa, S. K. (2015). Generalized log-logistic proportional hazard model with applications in survival analysis. *Journal of Statistical Distributions and Applications*, *3*(1). https://doi.org/10.1186/s40488-016-0054-z

Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine*, *4*(24). https://doi.org/10.21037/atm.2016.08.45