

# Parametric Survival Modeling of Tuberculosis Data -A Case Study of Federal Medical Centre, Bida, Nigeria

Vincent Abiodun Michael<sup>1</sup> & Ismaila Adewale Bolarinwa<sup>1</sup>

<sup>1</sup>Department of Statistics, The Federal Polytechnic, P.M.B. 55, Bida, Nigeria.

Correspondence: Vincent Abiodun Michael, Department of Statistics, The Federal Polytechnic, P.M.B. 55, Bida, Nigeria. E-mail: vicronz@gmail.com, iabolarinwa@gmail.com.

Received: March 3, 2020

Accepted: June 4, 2020

Online Published: June 10, 2020

doi:10.5539/mas.v14n7p37

URL: <https://doi.org/10.5539/mas.v14n7p37>

## Abstract

The research performed parametric survival analysis of Tuberculosis (TB) data (covering 2010 to 2016) collected from the Federal Medical Centre, Bida, Niger State, Nigeria. Three parametric survival models (Exponential, Weibull and Log-logistic) were fitted. The outcome variable was time to recovery from TB infection and four covariates being age, gender, TB type and occupation were involved. Models were estimated by maximum likelihood method and model selection criterion used was the Akaike Information Criterion (AIC). The exponential and log-logistic models found all covariates statistically insignificant while Weibull found all covariates but TB type significant at 5% level. Based on AIC, Weibull model with AIC of 163.5731 performed best, followed by log-logistic model with AIC of 191.419 and exponential model performed worst, with AIC of 517.9652. The best of fitted models being Weibull suggested that older patients had higher hazards than younger ones, older patients hence, had lower survival times, holding other covariates constant. That is, the older the TB patient, the lower was the time to recovery from TB. Males had higher hazards and hence, lower survival times compared to females. That is, male TB patients recovered faster than the females. Pulmonary TB patients had lower (insignificant) hazards and hence, higher survival times than Respiratory TB patients. TB patients on technical occupation had lower hazards than others and hence, had higher survival times than those whose occupations were considered not technical. The research concluded that age, gender and occupation were the major determinants of recovery period of TB patients. It was recommended that the Management of Federal Medical Centre, Bida, and other organizations involved in TB management could make use of the Weibull model to fit and predict both the survival and hazard rates of TB patients.

**Keywords:** Survival function, Hazard function, Tuberculosis, Exponential, Weibull, Log-logistic

## 1. Introduction

Survival analysis is the analysis of *time-to-event* data. Such data describe the length of time from a time origin to an endpoint of interest. Survival analysis methods are usually used to analyze data when major interest is *time* to occurrence of an *event*. Such event may be death, recovery, employment, marriage and so on. However, not all individuals could have experienced the event of interest before end of study, implying that real time of event is not available for some individuals, this is called *censoring*.

Survival models fall into parametric, non-parametric and semi-parametric schools of thought. The differences in the classes of models lie mainly in assumptions made regarding the distribution of the *survival time* (Kleinbaum & Klein, 2012). Kaplan Meier estimator readily comes to mind when discussing nonparametric approach. It has the shortcoming of being able to handle only one variable at a time. Cox proportional hazards model is an example of semi-parametric model.

Whenever survival times are assumed to conform to some specific statistical distribution, we talk of parametric survival model. Distributions that have been used for such include exponential, Weibull, gamma, generalized gamma, lognormal and log-logistic. Parametric models are not as commonly used in medicine as semi-parametric.

Survival models have been applied extensively in medical sciences, particularly in cancer studies. A few of such are Pocock, Clayton and Altman (2002), Bradburn, Love and Altman (2003a; 2003b) and Clark, Love and Altman (2003a; 2003b). Tuberculosis (TB) is however, of particular interest to this research. Walpola, Siskind,

Patel, Konstantinos and Derhy (2003) investigated TB prognostic factors in Australia. Ponnuraja and Venkatesan (2010) investigated TB cases in India; Fikadu, Yohannes and Fikre (2015) modeled risk factors of TB cases in Ethiopia. Xia, Ning and Huang (2018) performed the empirical comparison of the Breslow estimator and the KP estimator based on simulated data. Recent works include Yoosefi et al. (2018) who applied exponentiated Weibull model to cancer data, Odeleye, Bolarinwa and Bolarinwa (2019), an application of Kaplan Meier model on unemployment duration data, and Bolarinwa and Michael (2020) that fitted Cox proportional hazards model to Tuberculosis data. Khanum et al. (2019) applied logistic regression on TB study in Pakistan. Aung et al. (2019) applied Cox model to TB-HIV co-infected patients in Myanmar. Ogbo et al. (2018) observed a decreasing trend in TB disease burden among HIV-negative people in Nigeria.

TB is a widespread disease which can be fatal. It is infectious and caused by various strains of mycobacteria. It is spread through the air when people who have an active TB infection, cough, sneeze, or otherwise transmit respiratory fluids through the air. Most infections do not have symptoms, known as latent tuberculosis (Mason, Roy, Spillane & Singh, 2016).

Although a cure for TB was developed more than 50 years ago, TB remains one of the world's deadliest infectious diseases and is among the top ten causes of global morbidity and mortality (USAID, 2009). The increase in TB incidence is highest in Africa and Asia (World Health organization (WHO), 2009a; 2009b). This may be due to high poverty levels that both continents share in common. TB still constitutes a major health problem in Nigeria. This was recently brought to the fore by the WHO, which ranked Nigeria (fourth behind India, Indonesia and China) as one of the six countries which accounted for 60 per cent of the world total TB burden. Some 1.8 million people died from TB in 2015, 0.4 million of whom were co-infected with HIV (WHO, 2016).

The WHO 10-year strategy (2006-2015) to cut down the burden of TB in the world worked elsewhere as it reportedly saved some 37 million lives while some countries halved the prevalence of the disease. But in Nigeria, the reverse is the case. According to the National TB and Leprosy Control Programme (NTBLCP), over 600,000 new cases of TB occurred in Nigeria from a global report conducted in 2014. The fact that substantial numbers of the people infected in the country are unreported or undiagnosed, calls for concern as such aids spread of the disease. Most survival studies on TB have not incorporated occupation as a covariate, the present article does. Gender, TB type and age are other covariates in the model.

The main objective of this research is to model TB data using three parametric survival models: Exponential, Weibull and Log-logistic, with a view to selecting the best based on an information criterion.

## 2. Theoretical Framework and Model Specifications

Survival models are analyzed within the context of nonparametric, semi-parametric and fully parametric methods. The nonparametric approaches (Kaplan-Meier, Life-table and Cumulative hazard) only handle univariate cases, they cannot handle multivariate cases. Once the problem is multivariate in nature, there must be recourse to semi-parametric or fully parametric models which handle both univariate and multivariate survival cases.

In describing survival data, two essential functions to be considered are *survival function* (*survivor or survivorship function*) and *hazard function*.

Let  $T$  denote a continuous non-negative random variable representing survival time, with probability density function  $f(t)$  and cumulative distribution function  $F(t)$ .

The survival function,  $S(t)$  of an individual is the probability that the individual survives until at least time  $t$ .

$$S(t) = \Pr (T \geq t) \quad (1)$$

where  $t$  is a time of interest and  $T$  is the time variable.

The hazard function  $h(t)$  is a related measure of the probability that the event  $T$  occurs in the next  $(t + \delta t)$  given that the individual has reached time step  $t$ . It is defined

$$\begin{aligned} h(t) &= \lim_{\delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \delta t / T > t)}{\delta t} \\ &= \frac{f(t)}{S(t)} \end{aligned} \quad (2)$$

The hazard function is the rate of death or failure at an instant  $t$ , given that the individual survives up to time  $t$ . Hazard function is also known as instantaneous failure rate or the force of mortality (Angela, 2008; Singh & Mukhopadhyay, 2011).

When a model has covariates, survival modeling can be approached from four perspectives: Parametric families, accelerated failure time, proportional hazards and proportional odds. It is not all models that can be formulated using all approaches. However, some can be formulated using more than one approach. For example, exponential and Weibull models can be formulated as both accelerated failure time and proportional hazards models while the log-logistic model can be formulated as accelerated failure time model but not as proportional hazards model. It can however, be formulated as proportional odds model.

### 2.1 Parametric Survival Modeling Framework

Parametric modeling hinges on the principle that the outcome variable  $T$ , time to event, follows a specified probability distribution, whose *parameters*, are characteristically unknown and have to be estimated from available data. That is, they are fully parametric. Examples of these models include exponential model, Weibull model, and lognormal model (Richards, 2012). With a specified probability distribution of survival times, associated survival and hazard functions can be obtained.

For a specified survival times distribution  $f(t)$ , the survivor function  $S(t)$  is defined

$$S(t) = \int_t^{\infty} f(w)dw \quad (3)$$

and the hazard rate  $h(t)$  is

$$h(t) = \frac{f(t)}{S(t)} \quad (4)$$

The most commonly used parametric time-to-event models are the Weibull, log-logistic and log-normal distributions (Kahn & Khosa, 2016).

### 2.2 Proportional Hazards (PH) Modeling Framework

A very popular approach to modeling survival data is to relate survival times to explanatory variables (covariates) so that the effect of the covariates increases or decreases the hazard by a proportionate amount at all durations. That is, the effect of covariates is multiplicative with respect to the *hazard*  $h(t)$ . If we denote the average or hazard baseline function by  $h_0(t)$ , the hazard for a particular individual,  $h(t)$  is

$$h(t) = h_0(t)S(t) \quad (5)$$

where  $S(t)$  is usually written as

$$S(t) = \exp(\beta'X) \quad (6)$$

This effectively means that the hazard function  $h(t)$  of a certain individual is a multiple,  $S(t)$ , of the baseline hazard function,  $h_0(t)$ .

Taking log of (5), we have

$$\log[h(t)] = \log(h_0(t)) + \beta'X \quad (7)$$

where  $X$  is the set of covariates.

The function  $h_0(t)$  implies the underlying hazard for individuals with all covariates equal to zero. The fact that assumptions are not imposed on the distributional form of  $h_0(t)$  makes the PH modeling framework quite attractive. This explains why resulting models are described as *semi-parametric*.

From (7), it is evident that the PH model is a multiple linear regression of  $\log[h(t)]$  on the covariates with  $\log[h_0(t)]$  as the intercept. Proportionality in the PH model means that  $e^{\beta_i}$  are hazard ratios. A hazard ratio greater than 1 indicates that increase in the value of the variable is accompanied by increased event hazard and hence, decreased length of survival (Bradburn, Clark, Love & Altman, 2003).

Under proportionality assumption, the log of the hazard ratio for the  $i$ -th subject to the reference group is linear in the covariates. That is,

$$\log\left(\frac{h_i(t)}{h_0(t)}\right) = \beta' X \quad (8)$$

The Cox PH model is based on this framework and is the most commonly used for analyzing time to event data (Kleinbaum & Klein, 2012).

### 2.3 Accelerated Failure Rate (AFT) Models

This framework utilizes the relative length of time to the event rather than hazard or hazard ratio. The defining feature of AFT model is

$$S(t) = S_0(\varphi t) \quad (9)$$

where  $S_0(t)$  is the baseline function and  $\varphi$  is an *accelerator factor* (hence, the model name) that depends on the covariates according to the scheme

$$\varphi = \exp(\beta' X) \quad (10)$$

The scheme allows for evaluation of effect of covariates on survival time just as hazard ratio enables evaluation of covariates on the hazard. The idea is that each covariate either stretches or shrinks the survival curve along the time axis by a constant relative amount  $\varphi$ . So, the effect of the covariates is multiplicative with respect to *survival time*  $S(t)$ . If  $\varphi < 1$ , there is an *acceleration* (stretching) of the endpoint and if  $\varphi > 1$ , there is a *delay* (shrink). Acceleration factor  $\varphi$  is a ratio of survival times corresponding to any fixed value of  $S(t)$  (Kleinbaum & Klein, 2012). The median survival time for a group of individuals is  $\varphi$  times what they would have experienced in the reference group. Under the AFT framework, the survival time variable  $T$  is assumed to follow a specific distributional form.

The general form of AFT model is

$$\log(T_i) = \beta_{AFT} X_i + \sigma \mathcal{E} \quad (11)$$

where

$\log(T_i)$  is the log of survival time;

$\beta_{AFT}$  is the vector of AFT model parameters associated with covariate vector  $X_i$

$\mathcal{E}$  is a random error term;

$\sigma$  is a scale factor.

Parametric distributional assumption is usually imposed on the error term  $\mathcal{E}$ . Different probability distributions of  $\mathcal{E}$  and assumption on  $\sigma$ , generate different survival models. Commonly used distributions are exponential, gamma, extreme value, normal and logistic. The AFT formulation, (11) is a model of log-survival times as a linear function of the covariates,  $X_i$ .

Parameters under AFT and parameters under PH are related as follows:

$$\beta_{i AFT} = -\sigma \beta_i \quad (12)$$

### 2.4 The Exponential Model

This model is the simplest among the parametric survival models and it assumes constant, hazard over time. The distribution has only one parameter  $\lambda$  which is the mean and the variance. If the distribution times follow the exponential distribution,  $f(t)$  is defined

$$f(t) = \lambda e^{-\lambda t} \text{ for } t \geq 0 \quad (13)$$

Substituting (13) into (3), the survivorship function of the model is

$$S(t) = \exp(-\lambda t) \quad (14)$$

and substituting (13) into (4), the result is hazard function

$$h(t) = \lambda \quad (15)$$

Constant hazard means that the probability of experiencing the event remains same from one time interval to another- unrealistic for most modeling situations. The exponential model is clearly inadequate to model situations with varying hazard as its hazard is independent of time. A way out is to assume a density which results in variable hazard. For such densities, resulting hazards depend on time. The Weibull, log-normal and log-logistic readily come to mind.

Even when many possible options exist for relating the parameter  $\lambda$  to the vector of covariates  $\mathbf{X}$ , a natural choice is the log-linear model

$$\log(h) = \beta' X \tag{16}$$

The exponential model can be parameterized as both PH and AFT models.

To motivate exponential model as AFT model, we shall lean on (11). If  $\sigma = 1$  and  $\mathcal{E}$  is extreme value (min) distributed, the survival times  $T_i$  are exponentially distributed. As observed by Machin, Cheung and Parma (2005), the coefficients arising from PH and AFT formulations of the exponential model only differ in signs, implying that

$$\beta_{iAFT} = -\beta_i$$

This is evident from (12) when  $\sigma = 1$ , being a requirement of the exponential model. It is worthy of note that the two categories of coefficients have contrasting interpretations. That is, whereas a positive  $\beta_i$  implies that associated covariate  $x_i$  has a higher hazard and a shorter log survival time, same is communicated by a negative  $\beta_{iAFT}$ .

### 2.5 The Weibull Model

This is a modification of the exponential model. The modification is such that the survivorship and the hazard depend on time,  $T$ . The model is the most used of parametric survival models (Kleinbaum & Klein, 2012). If the distribution times follow the Weibull distribution,  $f(t)$  is defined

$$f(t) = \alpha \lambda t^{\alpha-1} e^{-\lambda t^\alpha} \tag{17}$$

If  $f(w)$  in (3) assumes the Weibull distribution, the survivorship function for the model is

$$S(t) = \exp[-(\lambda t)^\alpha] \tag{18}$$

and the hazard is

$$h(t) = \alpha \lambda (\lambda t)^{\alpha-1} \tag{19}$$

Parameter  $\alpha$  is a *shape* parameter, usually held fixed and it defines the shape of the hazard function. The presence of this shape parameter in the model offers a great flexibility.

If it is

- less than 1, hazard decreases over time;
- equal to 1, a constant hazard results and
- greater than 1, the hazard increases with time.

This range of shapes of possible hazard function makes the Weibull distribution attractive for modeling survival times (Machin, Cheung & Parmar, 2005). Parameter  $\lambda$  is the *scale* parameter and can be reparameterized in terms of the predictor variables (covariates). While holding  $\alpha$  fixed,  $\lambda$  can be formulated as done for exponential

$$\log(h) = \beta' X$$

It is worth mentioning that Weibull reduces to exponential model if  $\alpha = 1$ .

Leaning on the AFT model (11) once again, Weibull model results if the form of the error  $\mathcal{E}$  is extreme value distribution with  $\alpha = 1/\sigma$ . It is obvious that the model can be formulated as both PH and AFT models. Just like the exponential, the Weibull model is hence, a parametric PH model. It is the most used parametric model in survival analysis (Kleinbaum & Klein, 2012) and is the only distribution that is closed under PH and AFT

frameworks (Kalbfleisch & Prentice, 2002). It is in the literature (See Machin, Cheung & Parmar, 2005) that PH and AFT parameters for the Weibull are related as follows

$$\beta_{iAFT} = \frac{-\beta_i}{\alpha}$$

This can also be deduced from (12) as follows:

$$\beta_{iAFT} = -\sigma\beta_i$$

But for the Weibull model,  $\alpha=1/\sigma$ . Substituting  $1/\alpha$  for  $\sigma$ , we have

$$\beta_{iAFT} = \frac{-\beta_i}{\alpha} \quad (20)$$

### 2.6 The Log-logistic Model

The log-logistic model possesses survivorship function that varies with survival time. Unlike the exponential and Weibull models, log-logistic model can only be formulated as AFT model. It cannot be formulated as PH model but can be formulated as proportional odds model. Just like Weibull, this model has two parameters,  $\lambda$  and  $\gamma$  being scale and shape parameters respectively.

To motivate the model as AFT, we assume that the error term  $\varepsilon$  in (11) has a standard logistic distribution and  $\gamma = \frac{\sigma}{\sigma}$ . The survival times  $T_i$  are then log-logistically distributed, yielding log-logistic model.

The survivor function of the model is expressed as

$$S(t) = \frac{1}{(1+(\lambda t)^\gamma)} \quad (21)$$

and the hazard function is

$$h(t) = \frac{\lambda\gamma t^{\gamma-1}}{1+(\lambda t)^\gamma} \quad (22)$$

Parameter  $\gamma$  is the shape parameter; if it is less than or equal to 1, the hazard decreases continuously with time; and if it is greater than 1, the hazard first increases to a maximum point, then decreases with time. The hazard function is hence, *unimodal*.

## 3. Methodology

### 3.1 Data

The data was collected from Federal Medical Centre, Bida, Niger State, North Central Nigeria. The natives of Bida are by tribe Nupes. The data used for this study was secondary data collected from 259 TB patients' records covering 2010 to 2016. The covariates of interest in this study were age, gender, type of TB, and occupation. They were coded as follows:

Gender: Male = 1, Female = 0

Type of TB: Pulmonary TB (PTB) = 1, Respiratory TB (TB) = 0

Occupation: Technical = 1, Others = 0

Recovery: Alive = 1, Dead = 0

All occupations that typically expose the individual to chemicals or any other toxic substance were categorized as Technical.

### 3.2 Model Estimation

All models were estimated by the maximum likelihood method. To motivate discussion of maximum likelihood for survival models, recognition must be given to the fact that censoring is a feature that distinguishes survival data from other types of data. This special nature of survival data imparts on the derivation of the likelihood.

For survival data, a subject either experiences the event (fails) or is censored. The likelihood function is a combination of the likelihood for subjects that fail and those censored. If we assume right censoring, we have:

For a subject censored at  $t_i$

$$L_i(\beta) = S(t_i) = \exp\left[-\int_0^{t_i} h_i(u) du\right] \tag{23}$$

For a subject that fails at  $t_i$

$$L_i(\beta) = S(t_i)h(t_i) = h(t_i) \exp\left[-\int_0^{t_i} h_i(u) du\right] \tag{24}$$

Effectively, every subject contributes  $S(t_i)$  but only those who fail (i.e. experience the event) contribute  $h(t_i)$ .

Hence, the likelihood  $[L(\beta)]$  for all subjects is the product over all subjects.

$$L(\beta) = \prod_{i=1}^n h_i(t_i)^{\delta_i} \exp\left[-\int_0^{t_i} h_i(u) du\right] \tag{25}$$

where

$$\delta_i = \begin{cases} 0 & \text{if } i\text{-th subject is censored at time } t_i \\ 1 & \text{if } i\text{-th subject fails at time } t_i \end{cases}$$

Taking log of (25), we have

$$\text{Log}L = \sum_i [\delta_i \log h_i(t_i)] - \sum_i \Lambda_i(t_i) \tag{26}$$

where

$h_i(t_i)$  is hazard for subject  $i$  who fails at time  $t_i$  and

$\Lambda_i(t_i)$  [which is  $-\log S(t)$ ] is the cumulative hazard for subject  $i$  at his failure or censoring time.

Using the exponential model as an illustration, the procedure for estimation follows:

$$h(t) = \lambda; S(t) = \exp(-\lambda t) \text{ and } \Lambda(t) = \lambda t$$

Hence, for this model, (26) becomes

$$\text{Log} L = \sum_i [\delta_i \log h_i(t_i)] - \sum_i \lambda t_i \tag{27}$$

But  $\lambda = \exp(\beta' X)$

Substituting for  $\lambda$  in (27), we have

$$\text{Log}L = \sum_i [\delta_i \beta' X] - \sum_i t_i \exp(\beta' X) \tag{28}$$

On differentiating (28) with respect to  $\beta_0$ , we have the score equation

$$\frac{\partial \log L}{\partial \beta_0} = \sum_{i=1}^n [\delta_i - t_i \exp(\beta' X)]$$

and for the other parameters  $\beta_k, k=1, 2, \dots, p$ , the score equations are

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^n [\delta_i X_k - t_i X_k \exp(\beta' X)]$$

On setting the score equations to zero, MLE of  $\beta$ 's are obtained.

Leaning on (26), log of likelihood for the Weibull distribution is

$$\text{Log}L = \sum_i [\delta_i \log(\alpha \lambda t^{\alpha-1})] - \sum_i \lambda t_i^\alpha$$

and for the log-logistic, it is

$$\text{Log} L = \sum_i [\delta_i \log\left(\frac{\lambda \gamma t^{\gamma-1}}{1 + (\lambda t)^\gamma}\right)] - \sum_i -\log\left(\frac{1}{1 + (\lambda t)^\gamma}\right)$$

### 3.3 Models Comparison

Competing models can be assessed in a number of ways. The likelihood ratio (LR) test can be utilized for assessing models with a view to selecting the best. It does this by assessing whether additional parameter improves fit or not. It is however, applicable only when competing models are nested. Exponential model is nested within Weibull as the former is a particular case of the latter (when  $\alpha=1$ ). However, the log-logistic is not nested within either of the two models. This makes the comparison of the three models using the LR test inappropriate. Whenever such situation arises, the ideal is to resort to the use of an *information criterion*. An information criterion is useful for comparing both nested and non-nested models. It is a relative measure only as it does not inform about the quality of models, it only ranks the models being compared in order of goodness of fit. A number of such criteria exist. But the model selection criterion adopted in this article was Akaike information criterion (AIC).

The AIC proposed by Akaike (1974) is a measure of the goodness of fit of an estimated statistical model. AIC is given by

$$AIC = -2 \log(\text{likelihood}) + 2k \tag{29}$$

where  $k$  is the number of model parameters. Lower AIC indicates better likelihood. That is, lower AIC values indicate a better model. AIC is probably the most used of information criteria.

## 4. Results and Discussion

The data consisted of 259 registered TB patients treated between January 2010 to December 2016 at the Federal Medical Centre, Bida. Out of these patients, 148 representing 57.1% were males while 111 representing 42.9% were females. Sixty-six percent of the patients had Pulmonary TB while 34% had Respiratory TB. Patients aged between 0 – 20 years constituted 34.7% , those aged between 21 – 40 years were 30.9% while those aged between 41 – 60 years constituted 24.7% of the total. Patients aged 61 years and above constituted 9.7%. This suggests that the most youthful of the age groups (0-20 years) recorded highest incidence. Those on jobs regarded as technical constituted 59.1%.

Results of analysis are presented in Tables 1 to 3. Table 1 presents the parameter estimates, the corresponding hazard ratios and associated standard errors for each of the three models involved.

Table 1. A comparison of parameter estimates of Exponential, Weibull and Log-logistic models for time to recovery from TB

Covariates	Exponential			Weibull			Log-logistic		
	Estimate	H.R	S.E	Estimate	H.R	S.E	Estimate	H.R	S.E
Intercept	0.2673	1.306	0.1726	0.1549	1.168	0.0476	0.0112	1.011	0.0102
Age	0.0040	1.004	0.0036	0.0026	1.003	0.0011	0.0002	1.000	0.0002
Sex	-0.0755	0.927	0.1544	0.0881	1.092	0.0436	-0.0001	0.999	0.0092
Type	0.0495	1.051	0.1555	-0.0378	0.963	0.0417	0.0033	1.003	0.0003
Occupation	0.0492	1.050	0.1700	-0.1175	0.889	0.0470	0.0001	1.000	0.0099

Results of the exponential model (See Table 1) suggest that for each additional year to age of TB patient, the hazard rate (HR) increased by 0.4% (H.R = 1.004) holding the effects of other covariates constant, suggesting lower survival times for older patients. That is older patients recover faster than younger ones. With hazard rate of 0.927 for gender, male patients had 7.3% hazard rate lower compared to females, implying that males had



higher survival times than females, other covariates being constant, implying lower recovery period for females. Hazard rate of 1.051 for TB type implies that PTB patients recorded 0.51% hazard rate higher than patients in RTB category, holding other covariates constant. Hence, PTB patients had reduced survival times compared to RTB patients. Occupation with hazard rate of 1.050 implies that patients on technical jobs had 5% hazard rate higher than others, holding the effects of other covariates constant. This simply means that patients on technical occupation had reduced survival times when compared to patients on other occupations. That is, those on technical occupation recover faster than those on other occupations. It is worth noting that all hazard ratios of this model are not statistically significant at 5% level (See Table 2).

For the Weibull model (See Table 1) for a year increment in age of patients, the hazard rate increased by 0.3%, holding the effects of other covariates constant, thereby suggesting lower survival times for older patients. That is, lower times to recovery from TB for older patients compared to younger patients. This is a deviation from the expected as previous studies suggested otherwise. This may not be unconnected with the youthful nature of the subjects. Whereas those aged 0-40 years constituted 65.6% of the subjects involved in the study, only 9.7% were aged 61 years and above. This domination by the youths could be responsible for this inconsistent result. Age was recognized as one of major risk factors with survival of TB patients in several results in the past. Borgdorff, Buu, Houben, Qu, Lan and Cobelens (2007) identified age greater than 65 years as one of risk factors in survival of TB patients in Netherlands. Gopi., Vasantha and Subramani (2008) identified age as a risk factor among TB patients in South India. Pardeshi (2009) identified TB patients above 40 years as recording higher death rates than those younger. Balaky, Mawlood and Shabila (2019) observed lower survival rates among TB patients aged 65 years and above. They related this to possible presence of age-related diseases like cardiovascular diseases, diabetes mellitus and malignancy. Tolosie and Sharma (2014) observed that TB patients had statistically significant differences in survival experience due to age among other factors. But Ajagbe, Kabair and O'connor (2014) concluded no significant difference in survival experience of male and female and among age groups. In these studies, event was death.

HR of 1.092 for gender suggests 9.2% higher HR for male than for female TB patients. This implies lower survival times for males compared to female TB patients. That is, it took male TB patients shorter periods to recover than females holding other covariates constant. Balaky, Mawlood and Shabia (2019) observed higher, though not significant survival rates among male than female TB patients. Impact of gender on survival has presented a number of inconsistent results. Feng et al. (2012) in a study in Taiwan observed that more than three-fourths of TB patients were males and generally older than the females. According to Feng et al., (2012) higher mortality has been reported for males in Mexico, India and Italy. Limenih and Workie (2019) identified gender as one of the risk factors in their study of survival times to cure for TB patients in Ethiopia. The study concluded male multidrug-resistant TB patients experiencing longer recovery times than females. A similar study by Brust, Carrara, Osbum and Padayatchi (2010) in South Africa corroborated this. Sacks et al. (1988), Borgdorf (2007) and Zawdie, Enqueselasie and Tesfay (2015) confirmed higher hazard rates for female TB patients in South Africa, Netherlands and Ethiopia respectively. It is worthy of note that event in these cases was death and not recovery. A Cox PH analysis by Bolarinwa and Michael (2020) observed lower HR for male TB patients than for females in a recent study in which event of interest was recovery. So result of Cox analysis of same data is consistent with that of Weibull for the covariate, gender.

PTB patients had lower HR compared to RTB patients ( $H.R = 0.963$ ). Hence, PTB patients had higher survival times than RTB patients. That is, it took PTB patients longer periods to recover from the disease than RTB patients. Tolosie and Sharma (2014) did not identify TB type among others as a survival risk factor in a study in Ethiopia. However, Domingos, Caiaffa, and Colosimo (2008) in a similar study in Brazil, Munoz-Sellart, Hargreaves, Gausi, Kwanjana and Salaniponi (2001) in Malawi, and Borgdorff, Veen, Kalisvaart, and Nagelkerke (1998) in Netherlands observed TB type as a significant factor for mortality of TB patients. Omari-Sasu, Owusu, Boateng and Sabogu (2016) found TB type among factors responsible for varying survival experience among TB patients in Ghana. Compared to the exponential model, Weibull model has made contrary suggestions as regards hazards for gender, type of TB and occupation.

TB patients on technical jobs had lower hazards compared to others. Occupation is a category that has not been captured as frequently as age and gender in survival studies in medical research. However, Akinyemi, Sholanke and Odimegwu (2018) in a study identified type of occupation of parents as a risk factor for both infant and child mortality in Nigeria. Saroj, Murthy, Kumar, Singh and Kumar (2019) found husband's/partner's occupation significant in a study on under-five child mortality in India. Bolarinwa and Michael (2020) identified occupation as a component of the best of fifteen fitted Cox PH models of Tuberculosis data.

According to the log-logistic model, HR was constant for all ages. This contrasts with the position of the other two models. HR of 0.999 for gender implies a lower hazard (by 0.01%) for male than for female patients. The reduced hazard for males is in harmony with results obtained for the exponential model. HR of 1.003 for Type of TB implies 0.3% increase in the hazard for PTB compared for RTB patients. This is in agreement with the position of the exponential model. Type of occupation has a unit HR, suggesting constant risk for technical and other occupation. This contradicts the other two models which also maintain contrasting positions. It should be noted that parameters and hence, hazard ratios of this model are not statistically significant at 5% level (See Table 2), just like those of exponential model.

Table 2. P-values for Exponential, Weibull and Log-logistic model parameters

<i>Covariates</i>	<i>Exponential</i>	<i>Weibull</i>	<i>Log-logistic</i>
Intercept	0.121	0.001	0.274
Age	0.267	0.019	0.271
Gender	0.625	0.044	0.999
Type	0.750	0.365	0.722
Occupation	0.772	0.012	0.999

Table 2 presents the significance values (p-values) of parameters of each of the three models under consideration. The exponential model found all covariates insignificant at 5% level. This should be suggestive of a poor fit. High reported p-values suggest high level insignificance of model parameters. This means that going by this model, none of the covariates had a statistically significant influence on survival (time to recovery) of TB patients. The Weibull model found all covariates but type of TB significant at 5% level. Having three of four parameters significant may be indicative of a good fit. As far as this model is concerned, age, gender and occupation of TB patient exerted statistically significant influence on recovery from TB disease. The performance of the log-logistic model was similar to that of the exponential model as all covariates were found insignificant at 5% level. The Weibull model has performed better than the other two on the basis of proportion of significant parameters and hence, deserves better attention than its competitors.

Table 3. AIC in the analysis of time to recovery from TB

<i>Model</i>	<i>AIC</i>
Exponential	517.9652
Weibull	163.5731
Log-logistic	191.419

The AIC for each model is presented in Table 3. The Weibull model has the least AIC and has hence, performed best, followed by log-logistic and the exponential model performed worst. The performance of the Weibull above others based on AIC could not be attributable to chance but rather a corroboration of results of significance testing of models parameters presented in Table 2. This is consistent with results for Chere and Tesfay (2015) and Saroj, Murthy, Kumar, Singh and Kumar (2019) in which Weibull outperformed exponential, lognormal and log-logistic models based on AIC for a study on TB data in Ethiopia and India respectively. Little wonder that this model is most used of all parametric models in analyzing survival data.

This study is not without its limitations. Identified limitations follow: The data used is secondary and bias cannot be ruled out; among TB patients, time to recovery may be associated with presence of underlying illnesses which are not incorporated into the study; and age distribution of the subjects involved in the study is not necessarily representative of typical situation in the society.

**5. Conclusion**

This study has analyzed times to recovery of TB patients using the exponential, Weibull and log-logistic models. The Weibull model which found age, gender and occupation significant, outperformed exponential and log-logistic models on the basis of AIC. Weibull model hence, suggests that age, gender and occupation were the major determinants of recovery period of TB patients. The Management of Federal Medical Centre, Bida, Nigeria and other organizations involved in TB management could make use of Weibull model to fit and predict both the survival and hazard rates of TB patients.

## Acknowledgement

We acknowledge the editors of the Journal for their very useful comments and suggestions. The authors are grateful to TETFUND for financial support, the Management of Federal Medical Centre Bida for the data and the Management of The Federal Polytechnic, Bida for facilitating access to fund.

## References

- Ajagbe, O.B., Kabair, Z. & O'connor, T. (2014). Survival analysis of adult Tuberculosis disease. *PLOS One*, 9(11), 112838. <http://doi.org/10.1371/journal.pone.0112838>.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Akinyemi, J.O., Sholanke, B.L. & Odimegwu, C.O. (2018). Maternal employment and child survival during the era of sustainable development goals: Insights from proportional hazards modelling of Nigeria birth history data. *Annals of Global Health*, 84(1), 15-30. <http://doi.org/10.29024/aogh.11>.
- Angella, M.C. (2008). *Comparison between Cox proportional hazards and Weibull models*. Unpublished M.Sc. Thesis, Kansas State University.
- Aung, Z.Z., Saw, Y.M., Saw, T.N., Oo, N., Aye, H.N.N., Aung, S., Oo, H.N., Cho, S.U., Khaing, M., Kariya, T., Yamamoto, E. & Hamajima, N. (2019). Survival rate and mortality risk factors among TB-HIV co-infected patients at an HIV specialist hospital in Myanmar: A 12-year retrospective follow-up study. *IntJ Infec. Dis*, 80, 10-15.
- Balaky, S.T.J., Mawlood, A.A. & Shabila, N.P. (2019). Survival analysis of patients with Tuberculosis in Erbil, Iraqi Kurdistan region. *BMC Infec. Dis*, 19, 865. <https://doi.org/10.1186/s12879-019-4544-8>.
- Bolarinwa, I.A. & Michael, V.A. (2020). Survival modeling of Tuberculosis data- A case study of Federal Medical Centre, Bida, Nigeria. *Modern Applied Science*, 14(4), 99-110. <http://doi.org/10.5539/mas.v14n4p99>.
- Borgdorff, M. W., Buu, T. N., Houben, J., Quy, H. T., Lan, N. T. N. & Cobelens, F. G. J. (2007). HIV and Tuberculosis in Ho Chi Minh City, Vietnam, 1997–2002. *The International Journal of Tuberculosis and Lung Disease*, 13(10), 45–51.
- Bradburn, M.J., Clark, T.G., Love, S.B. & Altman, D.G. (2003a). Survival analysis Part II: Multivariate data analysis- an introduction to concepts and methods. *British Journal of Cancer*, 89, 431-436.
- Bradburn, M.J., Clark, T.G., Love, S.B. & Altman, D.G. (2003b). Survival analysis Part III: Multivariate data analysis- choosing a model. and assessing its adequacy of fit. *British Journal of Cancer*, 89, 605-611.
- Brust, J.C., Carrara, H. Osbum, G. & Padayatchi, N. (2010). High treatment failure and default rates for patients with multidrug-resistant Tuberculosis in Kwa-Zulu-Natal, South Africa, 2000-2003. *Int J. Tuber Lung Dis*, 14(4), 413-419.
- Chere, F.Z. & Enquoselassie, F. (2015). Parametric Survival Modelling of Risk Factor of Tuberculosis Patients under DOTS Program at Hawassa Town, Ethiopia. *International Journal of Biomedical and Clinical Engineering*, 4(1), 1-17.
- Clark, T.G., Bradburn, M.J., Love, S.B. & D.G. Altman, D.G. (2003a). Survival analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*, 36, 1221-1226.
- Clark, T.G., Bradburn, M.J., Love, S.B. & Altman, D.G.(2003b). Survival analysis Part IV: further concepts and methods in survival analysis. *British Journal of Cancer*, 89, 781-786.
- Domingos, M.P., Caiaffa, W.T. & Colosimo, E.A. (2008). Mortality, TB/HIV co-infection, and treatment dropout: predictors of Tuberculosis prognosis in Recife, Pernambuco State, Brazil. *Cadernos de Saude Publica*, 24(4), 887-896.
- Feng, J.Y., Huang, S.F., Ting, W.Y., Chen, Y.C., Lin, Y.Y., Huang, R.M., Lin, C.H., Hwang, J.J., Lee, J.J., Yu, M.C., Lee, Y.C.&Su, W.J. (2012). Gender differences in treatment outcomes of TB patients in Taiwan: a prospective observational study. *Clinical Microbiology and Infection*, 18(9), 331-337.
- Fikadu, Z.C., Yohannes, Y.T.& Fikre, E.(2015). Parametric survival modelling of risk factor of Tuberculosis patients under DOTS program at Hawassa town, Ethiopia. *International Journal of Biomedical and Clinical Engineering*, 4(1), 1-17. <http://doi.org/10.4018/UBCE.2015010101>.
- Gopi, P. G., Vasantha, M. & Subramani, R. (2008). Survival of Tuberculosis patients treated under DOTS in a

- Rural Tuberculosis Unit (TU), South India. *The Indian Journal of Tuberculosis*, 55, 64–69. PMID: 18516821.
- Kalbfleisch, J.D. & Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*. New Jersey: John Wiley.
- Khan, S.A. & Khosa, S.K. (2016). Generalized log-logistic proportional hazards model with applications in survival analysis. *Journal of Statistical Distributions and Applications*, 3(16). [Http://doi.org/10.1186/s40488-016-0054-z](http://doi.org/10.1186/s40488-016-0054-z).
- Khanum, Z., Jafri, Y.Z., Marri, A., Yasmeen, S., Malik, S., Raza, M. & Ulla, S.F. (2019). Prevalence and statistical analysis of various risk factors of Tuberculosis in Balochistan, Pakistan. *Int J. of Statistics and Applications*, 9(6), 191-200. <http://doi.org/10.5923/j.statistics.20190906.04>.
- Kleinbaum, D. G. & Klein, M.(2012).*Survival Analysis: A Self-Learning Text*(3rd ed.). New York: Springer.
- Limenih, Y.E. & Workie, D.L. (2019). Survival analysis of time to cure on multidrug resistance Tuberculosis patients in Amhara Region, Ethiopia. *BMC Public Health*, 19, 165. <https://doi.org/10.1186/s12889-019-6500-3>.
- Machin, D., Cheung, Y.B. & Parma, M.K.B.(2005). *Survival Analysis: A Practical Approach* (2nd ed.). Wiltshire: John Wiley.
- Mason, P.H. Roy, A., Spillane, J. & Singh, P. (2016). Social, historical and cultural dimensions of Tuberculosis.*Journal of Biosocial Science*, 48(2), 206-232. <http://doi.org/10.1017/S0021932015000115>. PMID 25997539.
- Odeleye, R.F.,Bolarinwa, B.T. & Bolarinwa, I.A. (2019). Survival modeling of unemployment duration experience of staff of National Bureau of Statistics, Ibadan, Nigeria. *International Journal of Innovative Research & Development*, 8(8), 29-32.<http://doi.org/10.24940/ijird/2019/v8/i8/JUL19016>.
- Ogbo, F.A., Ogeleka, P., Okoro, A., Olusanya, B.O., Olusanya, J., Ifegwu, I.K., Awosemo, A.O., Eastwood, J. & Page, A. (2018).Tuberculosis disease burden and attributable risk factors in Nigeria, 1990-2016. *Tropical Medicine and Health*, 46, 2-11. <https://doi.org/10.1186/s41182-018-0114-9>.
- Omari-Sasu, A.Y., Owusu, D.A., Boanteng, M.A. & Zabogu, Z. (2016). Survival analysis of Tuberculosis patients in Upper West Region of Ghana. *Int J of Stat&Apps*, 6(2), 40-44. <http://doi.org/10.5923/j.statistics.20160602.02>.
- Pardeshi, G. (2009). Survival analysis and risk factors for death in Tuberculosis patients on directly observed treatment-short course. *Int J Med Sc*, 63(5), 180-186. <http://doi.org/10.4103/0019-5359.53163>.
- Pocock, S.J., Clayton, T.C. & Altman, D.G. (2002). Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet*, 359, 1686-1689.
- Ponnuraja, C. & Venkatesan, P.(2010). Survival models for exploring tuberculosis clinical trial data—an empirical comparison. *Indian Journal of Science and Technology*, 2(7), 755-758.
- Richards, S. J. (2012). A handbook of parametric survival models for actuarial use. *Scandinavian Actuarial Journal*, 4, 233–257. <http://doi.org/10.1080/03461238.2010.506688>.
- Sacks, L. V. & Pendle, S. (1998). Factors related to In-Hospital deaths in patients with Tuberculosis. *Archives of Internal Medicine*, 158(17), 1916–1922. <http://doi.org/10.1001/archinte.158.17.1916> PMID: 9759688.
- Saroj, R.K., Murthy,K.H.,Kumar, M., Singh, R. & Kumar, A.(2019). Survival parametric models to estimate the factors of under-five child mortality. *J Health Res Rev*, 6, 82-88. [http://doi.org/10.4103/jhrr.jhrr\\_22\\_19](http://doi.org/10.4103/jhrr.jhrr_22_19).
- Singh, R. & Mukhopadhyay, K. (2011). Survival analysis in clinical trials: Basics and must know areas. *Perspec. Clin. Res*,2(4), 145–148. <http://doi.org/10.4103/2229-3485.86872>.
- Tolosie, K. & Sharma, M.K. (2014). *Application of Cox proportional hazards model in case of Tuberculosis patients in selected Addis Ababa Health Centers, Ethiopia*.2014, 1-11. <http://dx.doi.org/10.555/2014/536976>.
- USAID (2009). *The Global Burden of Tuberculosis*. Retrieved from <http://www.usaid.gov/pophealth/id/tuberculosis/burden.html>.
- Walpolo, H. C., Siskind, V., Patel, A.M., Konstantinos, A. & P. Derhy, P.(2003). Tuberculosis-Related deaths in Queensland, Australia, 1989-1998: Characteristics and risk factors. *The International Journal of Tuberculosis and Lung Disease*, 7(8), 742–750. PMID: 12921150

- World Health Organization (2009a). *Tuberculosis control in the South East Asia Region: WHO Annual Report 2009*. Geneva: WHO.
- World Health Organization (2009b). *Global tuberculosis control 2009: Epidemiology, strategy, financing: WHO Annual Report 2009*. Geneva: WHO.
- World Health Organization(2016). *National Tuberculosis and Leprosy control in Nigeria. WHO Annual Report 2016*. Geneva: WHO.
- Xia, F., Ning, J. & Huang, X. (2018). Empirical comparison of the Breslow estimator and the Kalbfleisch Prentice estimator for survival functions. *J BiomBiostat,9*, 392-398. <http://doi.org/10.4172/2155-6180.1000392>
- Yoosefi, M. Baghestani, A.M., Khadembashi, N., Pourhoseingholi, M.A., Baghban, A.A. & Khosrovirad, A. (2018). Survival analysis of colorectal cancer patients using Exponentiated Weibull Distribution. *Int J Cancer Manag, 11*(3), 1-6. <http://doi.org/10.5812/ijcm.8686>.
- Zawdie, F., Enquesolasie, F. & Tesfay, Y.Y. (2015). Parametric survival modelling of risk factor of Tuberculosis patients under DOTS program at Hawassa Town, Ethiopia. *International Journal of Biomedical and Clinical Engineering, 4*(1), 1-17. <http://doi.org/10.4018/IJBCE.2015010101>.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).