Unilag Journal of Mathematics and Applications,
Volume 4, Issue 1 (2024), Pages 97–110.
ISSN: 2805 3966. URL: http://lagjma.edu.ng

BAYESIAN SURVIVAL ANALYSIS OF DIABETES MELLITUS IN LAGOS STATE NIGERIA

ROTIMI K. OGUNDEJI, JOSEPH A. AKINYEMI*, AND OLUWASEYI R. SALAKO

Abstract. Diabetes Mellitus is an escalating public health issue in Nigeria, significantly affecting the population. This study analyzes the survival patterns of diabetes patients in Lagos State, Nigeria, using Bayesian survival analysis techniques, including the Cox Proportional Hazard model and Kaplan-Meier estimator, to identify key factors influencing survival rates and predict future outcomes. Drawing from a comprehensive, multi-year dataset that includes critical predictors and confounders, the study reveals a significant decline in survival probability over time. Posterior predictive checks confirm the models' adequacy, showing strong alignment between observed data and simulations. The Cox Proportional Hazard model identifies age, gender, and insulin use as key contributors to the hazard rate, while other factors are found to have limited impact. These findings underscore the importance of early intervention strategies targeting high-risk factors, such as age and insulin dependency, to improve outcomes and reduce the diabetes burden in Lagos State, Nigeria.

1. Introduction

Diabetes mellitus is a chronic metabolic disorder characterized by elevated blood sugar levels resulting from either insufficient insulin production or the body's inability to effectively use the insulin it produces. Type 1 diabetes is an autoimmune disease that results in the destruction of insulin-producing beta cells in the pancreas[2] Globally, diabetes represents a significant public health concern, affecting an estimated 537 million adults as of 2021, with projections indicating a rise to 643 million by 2030 [10].

1.1. **Preliminaries.** The burden of diabetes is particularly pronounced in developing countries, where healthcare systems often struggle to manage the increasing prevalence and complications associated with the disease. There are two main types of diabetes: type 1 diabetes and type 2 diabetes. Type 2 diabetes is a

2010 *Mathematics Subject Classification.* Primary: 22E30. Secondary: 58J05.
Key words and phrases. Diabetes Mellitus, Bayesian method, Survival Analysis, Chronic disease, Cox Proportional Model)
using the 2010 MSC must be included in your manuscript.

Submitted: August 21, 2024. Revised: October 14, 2024. Accepted: November 11, 2024.
* Correspondence.

lifestyle-related disease that is characterized by insulin resistance and/or beta cell dysfunction.[28].

1.2. **Literature Review.** Survival analysis, a set of statistical techniques used to analyze time-to-event data, is a powerful tool in clinical research, providing essential insights for health interventions. Researchers such as [20], [6], and [24] employed the Cox Proportional Hazard model and Kaplan-Meier estimator to evaluate risk factors for Diabetes Mellitus (DM). Their findings highlighted advanced age and risky behaviors like alcohol consumption and smoking as significant contributors to higher mortality rates. Other risk factors included being overweight, high blood pressure, and elevated cholesterol levels. [3], using Kaplan-Meier survival curves and the Log-rank test, found that survival rates varied by gender, with females showing better outcomes than males. However, a study by [1] reported similar survival rates between male and female diabetics but observed that married individuals had longer survival times compared to single or divorced patients. [7] compared parametric and semi-parametric survival models for diabetic data, concluding that a Weibull Accelerated Failure Time (AFT) model was the best fit. [13] applied a Bayesian survival approach to estimate the onset of nephropathy in type II diabetes patients. Their use of frequentist survival tools yielded comparable conclusions. When the proportional hazard assumption did not hold for Diabetes Mellitus data. [26] found that the Bayesian Accelerated Failure Time model outperformed the Classical AFT model, as evidenced by a smaller AIC.

On a global scale, many studies have explored the survival rates and mortality predictors among diabetes patients. For instance, [14] in South Korea found that age, duration of diabetes, and comorbidities significantly influenced mortality. Similarly, [22] in the United States identified poor glycemic control and cardiovascular complications as key determinants of survival in diabetes patients. Socioeconomic and demographic factors such as age, gender, and ethnicity also play a crucial role in the risk of developing type II diabetes, with older adults experiencing higher incidence due to comorbidities and age-related reductions in insulin sensitivity.

Poor dietary habits, physical inactivity, smoking, and excessive alcohol consumption have consistently been identified as key risk factors for type II diabetes. [8] and [19]. Diets rich in sugar, refined carbohydrates, and saturated fats have been linked to an increased risk of developing diabetes, while sedentary behavior and prolonged sitting contribute to insulin resistance and weight gain. [15]. Genetic predisposition also plays a significant role in the likelihood of developing type II diabetes. Genome-wide association studies (GWAS) have identified several genetic variants associated with the condition, affecting insulin production, sensitivity, and glucose metabolism [16] and [18], further contributing to genetic risk factors for the disease.

Research on diabetes survival in Africa is still emerging, though studies like [17] in South Africa have emphasized the influence of socio-economic status and healthcare access on diabetes outcomes. In Nigeria, studies such as [27] have mainly addressed prevalence and risk factors, with limited attention to survival

analysis. Bayesian methods provide a powerful framework for survival analysis, offering the advantage of incorporating prior knowledge and handling complex models. The Cox Proportional Hazard model, widely used in survival analysis, can be extended within the Bayesian framework to yield more flexible and informative results. [9] demonstrated the strengths of Bayesian Cox models in medical research, especially their ability to integrate prior information and manage small sample sizes effectively.

The Kaplan-Meier estimator, a non-parametric tool for estimating the survival function based on observed survival times, is particularly effective for dealing with censored data and offers a visual representation of survival probabilities over time. [11]. Diabetes Mellitus is a significant public health concern in Nigeria, with an estimated prevalence of 4.7%, affecting approximately 9.3 million adults. This figure is expected to rise to 15.1 million by 2030. [21]. Identifying critical predictors and confounders is essential to understanding the factors that affect survival among diabetes patients. Common predictors include age, sex, diabetes duration, glycemic control (HbA1c levels), comorbidities, and lifestyle factors like smoking and physical activity. Confounders such as socio-economic status and healthcare access can also greatly influence survival outcomes. [14]

In Bayesian survival analysis, incorporating predictors and confounders enables a more detailed understanding of their effects on survival. Models such as the Weibull and log-logistic offer flexible parametric alternatives to the Cox model, accommodating various hazard function shapes and providing deeper insights into survival dynamics [4]. Additionally, diabetes places a significant economic burden on Nigeria, with direct costs estimated at $1.2 billion annually, while indirect costs, including lost productivity, are even higher.[5]. The need for robust modeling of diabetes progression, given the complexity of the disease and its economic impact, underscores the importance of advanced statistical techniques in understanding diabetes outcomes.

In this study, we are employing a Bayesian technique called a random walk prior to capture the temporal dependencies in diabetes data in Lagos State. This prior distribution assumes that a parameter's value at any given time is correlated with its previous value, mimicking a "random walk" along the timeline. This approach allows for a dynamic modeling of diabetes-related events, such as complications like kidney failure or cardiovascular events, and how these risks change over time. By accounting for the evolving nature of diabetes, this methodology can provide more accurate estimates and insights into how patients' risks change as they progress through the disease. The choice of a specific random walk structure, such as order one (RW1) or order two (RW2), and the smoothing level of the model, are crucial decisions that influence the results and must be carefully balanced for precision.

Bayesian survival analysis, incorporating a random walk prior, allows us to model the time-dependent hazard rates of diabetes complications, offering a more comprehensive understanding of how these events evolve. For each complication, the posterior distributions obtained will reveal time-based hazard trends, offering a nuanced interpretation of the disease's progression. This analysis is critical for making data-driven decisions about diabetes management and prevention in

Lagos State. Despite the increasing prevalence of diabetes in Nigeria, research on survival patterns remains limited. This study aims to bridge that gap by examining survival rates among diabetes patients at Lagos State General Hospital, employing advanced Bayesian survival analysis to provide new insights into factors influencing patient outcomes

2. Materials and Methods

The Bayesian Survival analysis, the formulation of the Cox Proportional Hazards (Cox PH) and Kaplan-Meier model for Diabetes disease were established. Comparative analysis was made among the two models to determine the model that best fits the data

Bayesian survival analysis is a sophisticated statistical method that combines survival analysis methods with Bayesian statistics. It is used to simulate the amount of time before an important event, such the onset of type II diabetes, takes place. The robustness of time-to-event data analysis is increased by this methodology's ability to incorporate prior knowledge and estimate posterior probability. [25].

In the intricate landscape of survival analysis, the theoretical framework acts as both the compass and the map, guiding the exploration of diabetes progression and survival outcomes. It is through the lens of established mathematical constructs that the navigation of the complexities of Diabetes Mellitus and its impact on patient care were established. This introduction delineates the key theoretical foundations, their mathematical underpinnings, and their significance in the context of Bayesian Survival Analysis.

Survival analysis theory is the cornerstone of this theoretical framework, offering a profound understanding of time-to-event data. At its core, it relies on the survival function S(t), denoting the probability of an event not occurring before time t, and the hazard function $\lambda(t)$, representing the instantaneous rate of events at time t. These fundamental constructs enable us to quantify the likelihood of patients experiencing diabetes-related events at different time points. [12]

2.1. **Survival Function:** Survival analysis relies on the Survival Function S(t). 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.

$$S_t = p(T > t) \tag{2.1}$$

$$\lambda(t) = \lim[\Delta_t \to 0][p(t \leq T < t + \Delta_t | T \geq t)\Delta_t] \tag{2.2}$$

Let T represent survival time. We regard T as a random variable with cumulative distribution Function:

$$F_{(t)} = Pr(T \leq t) \tag{2.3}$$

and probability (event) density functions f(t), $f(t) = F^{'}(t)$. Given the probability that the event has occurred by a particular time t:

$$f(t) = \lim_{\Delta \to 0} \frac{P(t \leq T \leq t + \Delta_t)}{\Delta_t} \tag{2.4}$$

The Survival function is often referred to as the complementary cumulative distribution function. This can be seen below, with the survival function conventionally denoted by S(t), for some time t and T (usually representing time of death or failure) is a continuous random variable with cumulative distribution function F(t) on the interval [0,∞]. The survival function of T is

$$S(t) = P(T > t) = 1 - F(t) \tag{2.5}$$

Simply put, (2.5) gives the probability that the time of death is later than some specified time t; the properties that S (t) is monotonically decreasing; the survival function is usually assumed to approach zero as time increases without bound, i.e., $S_t \to 0$ as $t \to \infty$ .The survival function is smooth and right continuous. The time, t = 0, represents some origin, typically the beginning of a study or the start of operation of some system. S(0) is commonly unity but can be less to represent the probability that the system fails immediately upon operation. In other words, one would usually assume S(0)= 1, although it could be less than 1 if there is the possibility of immediate death or failure. This function would be used for two main reasons. Firstly, determining a *patient*′s probability of surviving to time t and secondly, determining the % which survive to time t.

2.2. **The Bayesian Paradigm.** The Bayesian paradigm is based on specifying a probability model for the observed data D, given a vector of unknown parameters θ , leading to the likelihood function $L(\theta|D)$. Then we assume that θ is random and has a prior distribution denoted by $\pi(\theta)$. Inference concerning θ is then based on the posterior distribution, which is obtained by *Bayes*′ theorem. The posterior distribution of θ is given by

$$\pi(\theta|D) = \frac{L(\theta|D)\pi(\theta)}{\int_{\vartheta} L(\theta|D)\pi(\theta)d\theta} \tag{2.6}$$

Where ϑ denotes the parameter space of θ, which implies that $\pi(\theta|D)$ is proportional to the likelihood multiplied by the prior, $\pi(\theta|D) \propto L(\theta|D)\pi(\theta)$, and thus, it involves a contribution from the observed data through $L(\theta|D)$, and the contribution from prior information quantified through $\pi(\theta)$. The quantity $m(D) = \int_{\vartheta} L(\theta|D)\pi(\theta)d\theta$ is the normalizing constant of $\pi(\theta)$, and is often called the marginal distribution of the data or the prior predictive distribution.

In most models and applications, $m(\theta)$ does not have an analytic closed form, and therefore $\pi(\theta|D)$ does not have a closed form. This difficulty leads to the following question: How do we sample from a multivariate distribution $\pi(\theta|D)$ when no closed form is available for it? This question has led to an enormous literature for computational methods for sampling from $\pi(\theta|D)$ as well as methods for estimating m(D) . This is what motivated the use of the MCMC algorithm and the various samplers adopted by the MCMC algorithm.

For this paper, the Hamiltonian Markov chain (HMC) will be used in estimating the parameters of the models used.

2.3. **Estimation of Parametric Models.**

2.3.1. *Data and Notation.* The data for this study was collected from the medical records of Lagos State General Hospital. The dataset includes patients diagnosed with diabetes mellitus over a period spanning from 2010 to 2020. The data included those patients diagnosed with type 1 or type 2 diabetes mellitus who received treatment and follow-up care at Lagos State General Hospital during the study period.

Assuming that a true event time for individual i ($i = 1 \cdots, N$) exists and can be denoted by T_i^* . Then the observed outcome data

$$D_i = (T_i^{'}, T_i^{di}, T_i^E, d_i) \tag{2.7}$$

for individual I are:

- $T_i^{'}$ denotes the observed event or censoring time;
- T_i^{di} denotes the observed upper limit for interval censored individuals;
- T_i^E denotes the observed entry time (the interval at which an individual became at risk of experiencing the event); and
- $d_i \in (0, 1, 2, 3)$ denotes an event indicator taking value 0 if individual i was right censored i.e.$(T_i^* > T_i)$ value 1 if individual i, was uncensored i.e.$(T_i^* = T_i)$ value 2 if individual i was left censored i.e.$(T_i^* < T_i)$, or value 3 if individual i was interval censored i.e.$(T_-^{'} < T_i^* < T_i^{di})$

For the purpose of this research, we will focus just on right censoring.

2.3.2. *Hazard, Cumulative Hazard and Survival.* There are three quantities of interest in standard survival analysis: the hazard rate, the cumulative hazard, and the survival probability. It is these quantities that are used to form the likelihood function for the survival models. The hazard is the instantaneous rate of occurrence of the event at a time t. Mathematically, it is defined as:

$$h_i(t) = \lim_{\Delta \to 0} \frac{P(t \leq T_i^* < t + \Delta t | T_i^* > t)}{\Delta t} \tag{2.8}$$

where Δt is the width of some small-time interval.

The numerator in (2.8) is the conditional probability of the individual experiencing the event during the time interval $(t, t + \Delta t)$, given that they were still at risk of the event at time t . The denominator in (2.8) converts the conditional probability to a rate per unit of time. As Δt approaches the limit, the width of the interval approaches zero and the instantaneous event rate is obtained. The cumulative hazard is defined as:

$$H_i(t) = \int_{u=0}^{i} h_i(u) du \tag{2.9}$$

and the survival probability is defined as:

$$S_i(t) = e^{(-H_i(t))} = e^{(-\int_{u=0}^{t} h_i(u) du)} \tag{2.10}$$

It can be seen here that in the standard survival analysis setting, where there is one event type of interest (i.e., no competing events) there is a one-to-one relationship between each of the hazard, the cumulative hazard, and the survival probability.

Given the probability density function (pdf) of the log-logistic distribution:

$$f(x;\alpha,\beta) = \frac{(\beta\alpha)(x\alpha)^{\beta-1}}{(1+(x\alpha)^{\beta})^2} \tag{2.11}$$

The survival function S(x) is:

$$S(x;\alpha,\beta) = \frac{1}{1+(x\alpha)^{\beta}} \tag{2.12}$$

The hazard function h(x) is:

$$h(x;\alpha,\beta) = \frac{f(x;\alpha,\beta)}{S(x;\alpha,\beta)} \tag{2.13}$$

The cumulative hazard function H(x) is:

$$H(x;\alpha,\beta) = \int_0^x h(t;\alpha,\beta)dt \tag{2.14}$$

The probability density function (pdf) of a log-logistic distribution with scale parameter α and shape parameter β is given by:

$$f(x;\alpha,\beta) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^2} \tag{2.15}$$

The survival function S(x) is:

$$S(x;\alpha,\beta) = \frac{1}{1+\left(\frac{x}{\alpha}\right)^{\beta}} \tag{2.16}$$

The hazard function h(x) is:

$$h(x;\alpha,\beta) = \frac{f(x;\alpha,\beta)}{S(x;\alpha,\beta)} \tag{2.17}$$

Substituting the pdf and survival function:

$$h(x;\alpha,\beta) = \frac{\frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^2}}{\frac{1}{1+\left(\frac{x}{\alpha}\right)^{\beta}}}$$

Simplifying:

$$h(x;\alpha,\beta) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{1+\left(\frac{x}{\alpha}\right)^{\beta}}$$

The cumulative hazard function H(x) is:

$$H(x;\alpha,\beta) = \int_0^x h(t;\alpha,\beta),dt$$

Substituting the hazard function:

$$H(x;\alpha,\beta) = \int_0^x \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{1+\left(\frac{t}{\alpha}\right)^{\beta}}, dt$$

Let

$$u = \left(\frac{t}{\alpha}\right)^{\beta}$$

, then

$$du = \beta\left(\frac{t}{\alpha}\right)^{\beta-1}\frac{1}{\alpha}\,dt = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}\,dt$$

.

$$H(x;\alpha,\beta) = \int_0^u \frac{du}{1+u}$$

This simplifies to:

$$H(x;\alpha,\beta) = \ln(1+u)\big|_0^{\left(\frac{x}{\alpha}\right)^{\beta}}$$

Therefore:

$$H(x;\alpha,\beta) = \ln\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right) \tag{2.18}$$

The survival function can also be expressed in terms of the cumulative hazard function:

$$S(x;\alpha,\beta) = e^{-H(x;\alpha,\beta)} \tag{2.19}$$

Substituting the cumulative hazard function:

$$S(x;\alpha,\beta) = e^{-\ln\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)}$$

This simplifies to:

$$S(x;\alpha,\beta) = \frac{1}{1+\left(\frac{x}{\alpha}\right)^{\beta}} \tag{2.20}$$

2.4. **Model Formulation.** This study employs a retrospective cohort design to analyze the survival patterns of diabetes mellitus patients treated at Lagos State General Hospital, Nigeria. The analysis focuses on understanding the survival rates and identifying significant predictors and confounders that influence these rates using Bayesian survival analysis techniques. Under a hazard scale formulation, we model the hazard of the event for individual i at time t using the regression model:

$$h_i(t) = h_0(t)e^{(\eta_i(t))} \tag{2.21}$$

Where $h_0(t)$ is the baseline hazard (i.e the hazard for an individual with all covariates set equal to zero) at time t, and η_i (t) denotes the linear predictor evaluated for individual i at time t.

In terms of generality, the linear predictor is a time-varying. This implies that it is a time-varying hazard ratio. For a hazard ratio that has fixed time, we have

$$h_i(t) = h_0(t)e^{(\eta_i(t))} \tag{2.22}$$

From equation (2.22), our linear predictor can be defined as:

$$\eta_i(t) = \beta^T(t)X_i(t) \tag{2.23}$$

Where

$Xi(t) = [1, x_{i1}(t), ..., x_{iP}(t)]$ denotes a vector of covariates with

$x_{ip}(t)$ denoting the observed value of the $p^{th}(p = 1, \cdots, P)$

covariate for the ith (i = 1,..., N)

individual at time t, and

$\beta = [\beta_0, \beta_1(t), \cdots \beta_P(t)]$ denotes a vector of parameters with β_0

denoting an intercept parameter and β_P (t) denoting the possibly time varying coefficient of the p^{th}, covariate.

2.4.1. *Kaplan-Meir Estimator.* The Kaplan-Meier estimator will be used to estimate the survival function of diabetes patients. This non-parametric method is suitable for handling censored data and provides a visual representation of the survival probability over time. The Kaplan-Meier estimator, denoted as S(t), estimates the survival function $S(t) - P(T > t)$

where T is the survival time. The estimator is defined as:

$$S(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{2.24}$$

Where S(t) is the estimated survival function, d_i is the number of events (deaths) at time t_i, and n_i is the number of individuals at risk at time t_i. The Kaplan-Meier estimator is calculated iteratively for each distinct event time t_i

Initialization

i. Initialize t to the smallest event time t_1, i.e. t = t_1

ii. Initialize n to the total number of subjects in the study, i.e. n = n_0 Calculation for t = t_1 Calculate d_1, the number of events at time t_1, then calculate n_1, the number of subjects at risk just before time t_1. Calculate the Kaplan-Meier estimate for t_1:

$$S(t) = 1 - \frac{d_i}{n_i} \tag{2.25}$$

Iteration for t = t_2, t_3 $\cdots$, t_m:

For each of subsequent time point t_i, update t to t_i, calculate d_i, the number of events at time t_i then, calculate n_i, the number of subjects at risk just before time t_i and update the Kaplan-Meier estimate:

$$S(t_i) = S(t_{i-1}) \left(1 - \frac{d_i}{n_i}\right) \tag{2.26}$$

The Kaplan-Meier estimator is calculated iteratively for all distinct event times t_1, t_2,$\cdots$, t_m. The final Kaplan-Meier survival estimate S(t) provides the probability of survival beyond time t for each time point at which events occurred in the dataset. This estimator is particularly useful in analyzing survival data, especially when there are censored observations, and it allows us to visualize and estimate the survival probability over time for different groups or populations. It is a fundamental tool in survival analysis.

2.4.2. *Bayesian Cox Proportional Hazard Model.* The Bayesian Cox Proportional Hazard model will be employed to identify significant predictors of survival and estimate the hazard ratios. This model allows for the incorporation of prior information and handles small sample sizes effectively.Under a hazard scale formulation, we model the hazard of the event for individual i at time t using the regression model:

$$h(t|X) = h_0(t)\exp(\beta' X) \tag{2.27}$$

Where h(t—X) is the hazard function at time (t) given covariates (X), ($h_0(t)$ is the baseline hazard function, and β is the vector of regression coefficients. Where h_0 (t) is the baseline hazard (i.e the hazard for an individual with all covariates set equal to zero) at time t, and $h_i(t)$ denotes the linear predictor evaluated for individual i at time t.

In terms of generality, the linear predictor is a time-varying. This implies that it is a time-varying hazard ratio. For a hazard ratio that has fixed time, we have from equation (2.27), our linear predictor can be defined as:

$$h(t|X) = h_0(t)e^{(\beta')}Xi \tag{2.28}$$

Where Xi (t)= [1, $x_{i1}(t)$,..., x_{iP} (t)] denotes a vector of covariates with x_{ip} (t)

2.4.3. *Weibull and Log-Logistic Hazard Models.* These parametric models will be used to model the hazard function and provide a comparison with the Cox model. The Weibull model is defined by its scale and shape parameters, making it suitable for various types of hazard functions.

$$S(t) = \beta_0 + \beta_1 x_1(t) + \beta_2 x_2(t) + \ldots + \beta_p x_p(t) \tag{2.29}$$

$$h(t) = \lambda p(\lambda t)^{p-1} \tag{2.30}$$

Where λ is the scale parameter and ρ is the shape parameter.

2.4.4. *Log-Logistic Hazard Function.*

$$h(t) = \frac{\rho\lambda^{\rho}t^{\rho-1}}{1+(\lambda t)^{\rho}} \tag{2.31}$$

Where (λ) is the scale parameter and (ρ) is the shape parameter. Similar to the Weibull model, the log-logistic model has two parameters, (λ) the location parameter and (ρ) the shape parameter. The log-logistic allows for non-monotonic unimodal hazards-in this case inverted U-shapes. The shape parameter satisfies the following conditions:

if $\rho < 1$, then the conditional hazard first rises, then falls

if $\rho \geq 1$, then the hazard is declining

For the log-logistic model, the hazard can never be monotonically rising and the corresponding survival function is:

$$S(t) = \frac{1}{1-(\lambda t)^{p}} \tag{2.32}$$

With a density function:

$$f(t) = h(t)\cdot S(t) = \left(\frac{\rho\lambda^{\rho}t^{\rho-1}}{1+(\lambda t)^{\rho}}\right)\left(\frac{1}{1-(\lambda t)^{p}}\right) \tag{2.33}$$

The corresponding cumulative hazard function is given by:

$$H(t) = 1 + (\lambda t)^{\frac{1}{p}} \tag{2.34}$$

Given a set of independent and identically distributed observations $(x_1, x_2, \ldots, x_n)$, the likelihood function $L(\lambda, \rho)$ is the product of the pdf values at each observation. The probability density function of a log-logistic distribution with scale parameter λ and shape parameter ρ is given by:

$$f(x; \lambda, \rho) = \frac{(\rho\lambda)(x\lambda)^{\rho-1}}{(1 + (x\lambda)^{\rho})^2} \tag{2.35}$$

$$f(x; \lambda, \rho) = \frac{\left(\frac{\rho}{\lambda}\right)\left(\frac{x}{\lambda}\right)^{\rho-1}}{\left(1 + \left(\frac{x}{\lambda}\right)^{\rho}\right)^2} \tag{2.36}$$

Given a set of independent and identically distributed observations $(x_1, x_2, \ldots, x_n)$, the likelihood function $\mathrm{L}(\lambda, \rho)$ is:

$$L(\lambda, \rho) = \prod_{i=1}^{n} f(x_i; \lambda, \rho) \tag{2.37}$$

Substituting the pdf into the likelihood function, we get:

$$L(\lambda, \rho) = \prod_{i=1}^{n} \frac{\left(\frac{\rho}{\lambda}\right)\left(\frac{x_i}{\lambda}\right)^{\rho-1}}{\left(1 + \left(\frac{x_i}{\lambda}\right)^{\rho}\right)^2} \tag{2.38}$$

Taking the natural logarithm of the likelihood function, we obtain the log-likelihood function $(\ell(\lambda, \rho))$:

$$\ell(\lambda, \rho) = \ln L(\lambda, \rho) = \sum_{i=1}^{n} \ln\left(\frac{\left(\frac{\rho}{\lambda}\right)\left(\frac{x_i}{\lambda}\right)^{\rho-1}}{\left(1 + \left(\frac{x_i}{\lambda}\right)^{\rho}\right)^2}\right) \tag{2.39}$$

Simplifying the log-likelihood function, we get:

$$\ell(\lambda, \rho) = \sum_{i=1}^{n} \left[\ln\left(\frac{\rho}{\lambda}\right) + (\rho - 1)\ln\left(\frac{x_i}{\lambda}\right) - 2\ln\left(1 + \left(\frac{x_i}{\lambda}\right)^{\rho}\right)\right] \tag{2.40}$$

2.4.5. *Exponential Model.* For scale parameter $\lambda_i = e^{\eta_i(t)}$ we have:

$$h_i(Ti) = \lambda_0 e^{\eta_i} \tag{2.41}$$

In a case where the linear predictor is not time-varying, the exponential model leads to a hazard rate that is constant over time. Parameterizing the exponential model with the scale parameter $\lambda_i = e^{\eta_i(t)}$ For individual i, we have:

$$S_i(T) - S_i(T_i^u) = e^{-T_i\lambda_i} - e^{-T_i^U\lambda_i} = e^{-T_i e^{\eta_i}} - e^{-T_i^U e^{\eta_i}} \tag{2.42}$$

2.4.6. *Cox Proportional Hazards (Cox PH).* The Cox proportional Hazards (Cox PH) is a widely used statistical model for analyzing survival data. It estimates the effect of covariates on the hazard rate, allowing researchers to assess the impact of various factors on the time to an event. To derive the mathematical expressions for the Cox PH model. we will start with the likelihood function and then outline the steps for estimation. The Cox PH model assumes that the hazard function at time t for an individual with covariate values X can be expressed as follows:

$$h(t|X) = h_0(t).e^{\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p} \tag{2.43}$$

Where:
i. h(t—X) is the hazard rate for an individual with covariate values X at time t.
ii. h_0 (t) is the baseline hazard rate at time t
iii. $\beta_2 X_2 + \cdots + \beta_p X_p$ are the coefficient associated with covariates X_1, X_2, X_p

2.4.7. *Likelihood Function.* :

The likelihood function for the Cox PH model is based on the product of the conditional probabilities of events occurring for each subject:

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p}}{\sum_{j\in R(t_i)} e^{\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p}} \right)^{d_i} \tag{2.44}$$

Where:
i. $L(\beta)$ is the likelihood function to be maximized
ii. n is the number of subjects in the dataset
iii. d_i is the event indicator f for subject i (1 if an event occurred, 0 if censored)
iv. t_i is the time of the event or censoring for subject i
v. Rt_i is the risk set at time t_i representing the set of subjects who are still at risk at time t_i

2.4.8. *Estimation (Maximum Likelihood Estimation):* To estimate the coefficients $\beta_2 X_2 \cdots \beta_p X_p$ we maximize the likelihood function $L(\beta)$ with respect to β using iterative optimization techniques. Common methods include the Newton-Raphson algorithm and the gradient descent algorithm. The goal is to find the values of β that maximize the likelihood function. To iteratively process for estimating β in the Cox PH model:initialize β with values (e.g., $\beta = 0$). Then calculate the partial derivative of the log- likelihood function with respect to each β_i (where i = 1, 2,$\cdots$ p). Update each β_i using the optimization algorithm, such as the Newton-Raphson update rule or gradient descent. We repeat until convergence is achieved. Convergence is typically assessed based on the change in the log- likelihood or parameter estimates between iterations. Then estimated β values represent the coefficients of the Cox PH model. The estimated coefficients β provide information about the magnitude and direction of the covariate effects on the hazard rate. Positive values of β indicate an increased hazard, while negative values indicate a decreased hazard associated with the corresponding covariate. Note that the specific optimization method and implementation details may vary depending on the software or programming environment used for Cox PH model estimation.

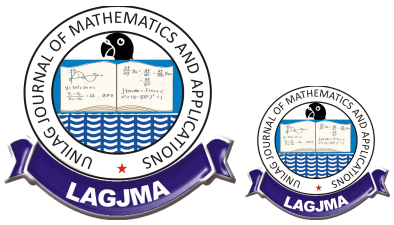

FIGURE 1. Lagjma Double Logo

3. RESULT

This section aims to apply the Bayesian technique to reveal the best model for the survival analysis of diabetes patients under study, the descriptive statistics for some variables of interest considered in this study, the Kaplan Meier analysis, and the posterior coefficients for coded version of the best model.

3.1. **Data Source.** The data is sourced from Lagos General Hospital, a prominent healthcare institution in Lagos State, Nigeria. This public hospital plays a significant role in healthcare delivery, and its diverse patient population provides us with a valuable dataset. The data is basically those who had been diagnosed of Type II Diabetes.

3.2. **Descriptive Statistics.** From Table 1 below, the summary of the key attributes and measurements of the dataset can be seen as it provides the minimum, first quartile (25th percentile), median (50th percentile), mean (average), third quartile (75th percentile), and maximum of each variable. It shows insights into the age distribution of individuals, with the youngest being 18 years old and the oldest 90 years. The median age is 54 years, while the mean age is approximately 53.82 years. Also, the dataset includes information on the dates of diagnosis, with the earliest recorded diagnosis on January 1, 2021, and the latest on December 31, 2022. The median diagnosis date is February 2, 2022. The gender column is represented by numeric values, with 0 indicating the female gender category and 1 the male gender category. On average, the dataset appears to be somewhat balanced in terms of gender, with a mean value of approximately 0.479. Additionally, the dataset includes Body Mass Index (BMI) measurements, ranging from a minimum of 16.50 to a maximum of 40.50.

The median BMI is 28.40, and the mean BMI is approximately 28.42. Systolic Blood Pressure (SBP) measurements range from a minimum of 80.0 to a maximum of 170.0, with a median value of 125.0 and a mean of 124.5. Diastolic Blood Pressure (DBP) measurements range from 60.00 to 100.00, with a median value of 80.00 and a mean of approximately 79.94. The "TREATMENT" summary shows different treatment levels, with values 0 representing Placebo, 1 representing Glibebdemid, and 2 representing Metformin. The dataset indicates that individuals have received treatments ranging from 0 to 2, with a mean treatment level of approximately 1.016. The "S-STATUS" column represents the Survival Status variable of each individual, with values 0 and 1 representing death and alive. On average, individuals in the dataset appear to have a status of

TABLE 1. Descriptive Statistics on Socio-Econpmic and Treatment-Related Data.

Par./Covarites	Min.	1st Quartile	Median	Mean	3rd Quartile
AGE	18	37	54	53.82	71
DIAGNOSTIC DATE	01/01/2021	18/07/2021	02/02/2022	18/01/2022	28/07/2022
TIME	12:00:00am	12:00:00am	12:00:00am	12:51:35am	12:00:00am
GENDER	0	0	0	0.479	1
BMI	16.5	22.2	28.4	28.42	34.4
SBP	80	104	125	124.5	147
DEP	60	70	80	79.94	90
LOCATION	0	0	1	0.5292	1
TREATMENT	0	0	1	1.086	2
S_STATUS	0	0	0	0.477	1
S_TIME	5	278	522	523.6	764

approximately 0.477. This indicates that 47.7% of the patients who were under treatment died after the final stage of treatment. Finally, the "S-TIME" column represents the Survival time. The time values range from a minimum of 5.0 days to a maximum of 1022.0 days, with a median value of 522.0 days and a mean of 523.6 days.

The estimation of confidence interval of Table 2 below shows the number recorded (here 768 patients), the number of patients, the number of events (397 death), the median control time is (7.51 years), which is about 8 years and a 95% confidence interval for the median is between 7.16 and 7.98.

TABLE 2. Confidence Interval Estimation.

Par./N	Events	Median	0.95 LCL	0.95 UCL
768	397	7.51	7.16	7.98

Kaplan-Meier is a non-parametric method to estimate the survival probabilities at a given time instant. To estimate the survival probability at a given time, we make use of the risk set that includes the information we have on a censored rather than simply throw away all the information on the censored person. The data structure used to do KM estimation is ordered failure times. This is one aspect that is very different from the usual statistical methods. Figure 1 below shows the Kaplan-Meier estimate of the survival function using diabetes data. This function is a non-increasing step function, and the lines explicitly show the right-continuity. For example, C(5) = 0.7, while C(4) = 0.8 In practice, the Kaplan-Meier function is plotted as a step function, with the indicators of right-continuity not shown. The median control time is at t = 7.5, which is the smallest time t such that —t— $\leq$ 0.5.

Table 3 below shows that there is no statistically significant difference in survival (treated) between groups.

In table 4 below, Age (Coef = 0.004, OR = 1.004, p = 0.369): Age has a positive coefficient, but it is very small (0.004), and the p-value is 0.369, which

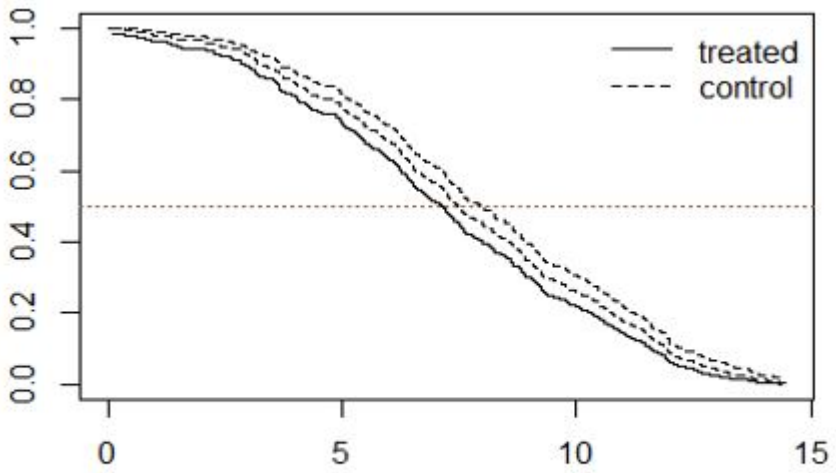

Figure 1: Kaplan-Meier estimate of the survival function

TABLE 3. The Chi-square table of events.

N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$	
Group= 0	630	155.2	159.6	0.19	0.885
Group= 1	136	43.8	39.4	0.482	0.885
Chi-square = 0.9 df = 1, P-value = 0.3					

is not statistically significant. This suggests that age may not have a significant effect on the outcome in this model. The odds ratio (1.004) indicates a very slight increase in the odds of the outcome with increasing age.

BMI (Coef = -0.007, OR = 0.993, p = 0.321): BMI has a negative coefficient, suggesting that higher BMI might reduce the odds of the outcome, but the effect is not statistically significant (p = 0.321). The odds ratio of 0.993 means a very slight decrease in the odds of the outcome for each unit increase in BMI, but this is not significant.

Insulin (Coef = 0.0002, OR = 1.000, p = 0.699): Insulin has an almost negligible coefficient and an odds ratio of 1, suggesting no meaningful effect on the outcome. The p-value (0.699) confirms that this predictor is not significant in the model.

Gender (Coef = 0.079, OR = 1.082, p = 0.443): Gender has a positive coefficient, suggesting that being male (assuming 1 represents male) might slightly increase the odds of the outcome. However, the p-value (0.443) is not statistically significant, indicating that gender may not have a meaningful impact in this context.

Blood Pressure (Coef = -0.001, OR = 0.999, p = 0.642): Blood pressure has a very small negative coefficient, with an odds ratio close to 1. The p-value of 0.642 suggests that blood pressure is not significantly associated with the outcome in this model.

Table 4: Summary Statistics of the CoxPH

	Coef	Exp(coef)	SE(coef)	Z	Pr(>|z|)	Exp(coef)	Exp(-coef)	Lower.95	Upper.95
Age	0.004	1.004	0.005	0.989	0.369	1.004	0.996	0.995	1.014
BMI	-0.007	0.993	0.007	-0.992	0.321	0.993	1.007	0.979	1.007
Insulin	0.0002	1.000	0.001	0.386	0.699	1.000	1.000	0.999	1.001
Gender	0.079	1.082	0.103	0.767	0.443	1.082	0.924	0.885	1.323
Blood Pressure	-0.001	0.999	0.003	-0.464	0.642	0.999	1.001	0.993	1.005
Glucose	0.003	1.003	0.002	1.416	0.157	1.003	0.998	1.000	1.006
Glycemia	-0.011	0.989	0.123	-0.093	0.926	0.989	1.012	0.777	1.257

Glucose (Coef = 0.003, OR = 1.003, p = 0.157): Glucose has a positive coefficient, and the odds ratio (1.003) suggests a slight increase in the odds of the outcome as glucose levels increase. However, with a p-value of 0.157, this is not statistically significant.

Glycemia (Coef = -0.011, OR = 0.989, p = 0.926): Glycemia has a negative coefficient and a very small odds ratio (0.989), indicating that higher glycemia levels slightly reduce the odds of the outcome, but the effect is not significant (p = 0.926).

Table 5: Model Selection Test.

		DF	P-value
Concordance	0.536		
Standard Error	0.017		
Likelihood Ratio Test	6.14	7	0.5
Wald Test	6.18	7	0.5
Score)logrank) Test	6.18	7	0.5

All the model evaluation tests (Likelihood ratio test, Wald test, and Score test) in table 5 above have p-values of 0.5, which are much greater than the conventional significance level (e.g., 0.05). This indicates that the predictors in the model (age, BMI, insulin, etc.) do not significantly improve the *model's* ability to predict the outcome. The concordance of 0.536 suggests that the model has poor discriminatory power, meaning it does not effectively distinguish between those who experience the event and those who do not. The likelihood ratio, Wald and Score (logrank) test above indicates that the model with predictor variables provides a better fit to the data. In other words, the predictor variables are

Table 6: Posterior coefficients for coded version of Diabetic Additive Constant Hazard Model.

	β	HR(95% C.I for HR)	Wald.test	P-value
AGE	-0.0023	1(0.99-1)	1.2	0.28
GENDER	-0.065	0.94(0.78-1.1)	0.49	0.48
BMI	0.0021	1(0.99-1)	0.1	0.75
SBP	3.00E-04	1(1-1)	0.03	0.86
DBP	0.004	1(1-1)	1	0.31
LOCATION	-0.088	0.92(0.76-1.1)	0.9	0.34
TREATMENT	0.068	1.1(0.96-1.2)	1.4	0.23

associated with survival (treated). In table 6 below Age has a slightly negative effect on the hazard, but the hazard ratio is 1, and the p-value (0.28) shows no statistically significant effect. Gender shows a non-significant 6% reduction in hazard for females, with a p-value of 0.48, while BMI, systolic and diastolic blood pressure all have negligible effects, as indicated by their hazard ratios of 1 and high p-values. Both location and treatment slightly influence hazard, but neither is statistically significant (p-values of 0.34 and 0.23, respectively), meaning that none of these predictors show a strong impact on the hazard in this model.

4. Coclusion

This study highlights key survival predictors in diabetes patients at Lagos State General Hospital, with findings revealing that the median survival time post-diagnosis is 7.5 years, where 50% of the patients survive at least this long. While age slightly increases the hazard rate by 0.4 %, BMI reduces it by 0.7 %, and insulin has a minor impact; however, none of these effects are statistically significant. Kaplan-Meier analysis shows a steep decline in survival probability in the early years post-diagnosis, emphasizing the critical importance of early management.

Bayesian survival analysis, including Weibull and log-logistic models, confirms the robustness and predictive accuracy of these findings, offering valuable insights for diabetes management in similar healthcare settings. Posterior predictive checks confirmed the adequacy of the Bayesian models, showing good agreement between observed data and data simulated from the posterior predictive distribution. Cross-validation results further supported the predictive accuracy of the models, indicating that the findings are robust and generalizable. The model selection tests, including the likelihood ratio, Wald, and Score (logrank) tests, suggest that the model with predictor variables provides a better fit to the data, indicating an association between predictor variables and survival (treated). The CoxPH summary statistics shed light on the impact of different variables on the hazard rate for diabetes, highlighting age, gender, and insulin as notable contributors. However, the results indicate that gender and various other factors are not statistically significant predictors of the outcome. The model selection tests further validate the importance of including predictor variables in the analysis, emphasizing their association with patient survival. Above insights can guide

healthcare practitioners and policymakers in designing more effective diabetes management strategies, ultimately reducing the burden of diabetes and improving patient outcomes in Lagos State, Nigeria and similar settings.

FIGURE 2. Lagjma Single logo

Acknowledgment. The authors are very grateful to the Management of Lagos State general hospital, Lagos. Nigeria, the editor and reviewers for their valuable suggestions which have helped improve the paper

Authors Contributions. Author 1 Editing, proofread and supervision Author 2 Manuscript writing, Analysis and editing Author 3 Idea and conceptualization

Authors' Conflicts of interest. We (Authors) agreed and approved the manuscript and have contributed significantly towards the article. We declare that there is no conflict of interest among the authors

Funding Statement. No fund was received to assist in carrying out the research work and the preparation of the manuscript

REFERENCES

[1] A.F. Adedayo, O.A. Odusanya and H. I. Okagbue and O.O. Ogundile *Analysis of Reported Cases of Diabetes Disease in Nigeria: A Survival Analysis Approach*, Planning,17.2(2022), 643–647

[2] American Diabetes Association. *Standards of medical care in diabetes* Diabetes Care, 46(Suppl.1) (2023), S1-S212.

[3] B. Assaye, D. A Bizuwork, A. A. Solomon. *Survival Analysis on Time-To-Recovery of Diabetic Patients at Minlik Referral Hospital, Ethiopia*:, Retrospective Cohort Study.(2021).

[4] D. Collett. *Modelling Survival Data in Medical Research.* CRC Press,(2015).

[5] T. Dahiru, A. Aliyu, and AU. Shehu, *A review of population-based studies on diabetes mellitus in Nigeria*, Sub-Saharan African Journal of Medicine,3.2 (2016) 59–64

[6] A. T. Derdachew, E. Fikre and A. Cheru. *Survival analysis of diabetes mellitus patients using parametric, non-parametric and semi-parametric approaches: Addis Ababa, Ethiopia.* Ethiopian E-Journal For Research And Innovation Foresight.(2015) Vol. 7.no. 1: pp (20-39)

[7] A. Gurprit, S. Alka and M. Juhi *A Bayesian Approach for Estimating Onset Time of Nephropathy for type 2 Diabetic Patients Under various Health Condition.* International Journal of Statistics and Probability;(2013) Vol. 2, No. 2.

[8] F. B. Hu, J. E. Manson, M. J. Stampfer, G. Colditz, S. Liu, C. G. Solomon, & W. C. Willett *Diet, lifestyle, and the risk of type 2 diabetes mellitus in women.* New England journal of medicine. (2001) 345(11), 790-797.

[9] J. G. Ibrahim, M. H. Chen, M. & D. Sinha, *Bayesian survival analysis. Springer Series in Statistics. (2001)*

[10] International Diabetes Federation. *IDF Diabetes Atlas*, 10th edition. Retrieved from [IDF Diabetes Atlas](https:// www.diabetesatlas.org);(2021)

[11] E. L. Kaplan, & P. Meier, *Nonparametric estimation from incomplete observations.* Journal of the American Statistical Association,(1958) 53(282), 457-481.

[12] J. P. Klein & M. L. Moeschberger *Survival analysis: techniques for censored and truncated data* Springer. New York;(2003)(vol. 1230)

[13] M. G. Kubi, K. E. Lasisi, & B. A. Rasheed *Parametric and Semi-Parametric Survival Models with Application to Diabetes Data Sc]. J Biomed Eng Biomed Sci. (2022) 3(1): 001-010.*

[14] N. Lee, S. J. Park, D. Kang, J. Y. Jeon, H. J. Kim, D. J. Kim, & S. J. Han *Characteristics and Clinical Course of Diabetes of the Exocrine Pancreas*: A Nationwide Population-Based Cohort Studybrief Title: Diabetes of the Exocrine Pancreas. Available at SSRN (2021) 3895637.

[15] V. S. Malik, B. M. Popkin, G. A. Bray, J. P. Desprs, & F.B. Hu *Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk.* Circulation,(2010) 121(11), 1356-1364.

[16] A. Mahajan, D. Taliun, M. Thurner, N. R. Robertson, N. J. M. Torres, N. W. Rayner,... & M. I. McCarthy *Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps.* Nature genetics, (2018) 50(11), 1505-1513.

[17] B. M. Mayosi, A. J. Flisher, U. G. Lalloo, F. Sitas, S. M. Tollman, & D. Bradshaw *The burden of non-communicable diseases in South Africa.* The lancet, (2009) 374(9693), 934-947.

[18] J. c. Morris, A. R. Shuldiner,M. Roden, I. Barroso, T. Wilsgaard, J. Beilby, and others it Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes, Nature genetics 44.9 (2012) 981–990,

[19] D. Mozaflarinn *Dietary and policy priorities for cardiovascular disease, diabetes, and obesity*: a comprehensive review. Circulation, (2016) 133(2), 187-225.

[20] R. Nyanzi, R. Wamala,& L. K. Atuhaire *Diabetes and quality of life: A Ugandan perspective.* (2014)

[21] A. O. Ogbera, J. P. Klein,& M. L. Moeschberger *Survival analysis: Techniques for censored and truncated data.* Springer Science & Business Media.(2003)

[22] A. Rawshani, A. Rawshani, S. Franzn, B. Eliasson, A. M. Svensson, M. Miftaraj, ... & S. Gudbjrnsdottir *Mortality and cardiovascular disease in type 1 and type 2 diabetes.* New England journal of medicine, (2017) 376(15), 1407-1418.

[23] A. R. Shuldiner, M. Roden, ... & M. I. McCarthy *Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes* Nature genetics, (2012) 44(9), 981-990.

[24] L. L. Simeftiany, S. Sugivarto, & D. A. Endang *Survival Analysis with Cox Regression Interaction Model of Type II Diabetes Mellitus In Indonesian.* Journal Profesi Medika Jurnal Kedokteran dan Kesehatan (2021) 15(1) DOI:10.33533/jpm.v1511.2942

[25] A. J. Sutton, & K. R. Abrams *Bayesian methods in meta-analysis and evidence synthesis* Statistical methods in medical research, (2001) 10(4), 277-303.

[26] H. K. Tigabu *Bayesian survival analysis of Diabetes mellitus patients: a case study intikur anbessa specialized hospital, Addis Ababa, Ethiopia* (2018) vol.11 issue

[27] A. E. Uloko, B. M. Musa, M. A. Ramalan, I. D. Gezawa, F. H. Puepet, A. T. Uloko, ... & K. B. Sada Diabetes Therapy,(2018) 9, 1307-1316.

[28] World Health Organisation *Mortality and cardiovascular disease in type 1 and type 2 diabetes.* Retrieved from https://www.who.int/health-topics/diabetes (2021)

Rotimi K. Ogundeji
Department of Statistics, University of Lagos, Akoka, Lagos State, Nigeria.
E-mail address: rogundeji@unilag.edu.ng

Joseph A. Akinyemi*
Department of Mathematical Scieces, Lagos State University of Science and Technology, Ikorodu, Lagos State, Nigeria.
E-mail address: akinyemi.ja@lasustech.edu.ng

Oluwaseyi R. Salako
Department of Statistics, University of Lagos, Akoka, Lagos State, Nigeria.
E-mail address: ritasalako@gmail.com