Analysis of Reported Cases of Diabetes Disease in Nigeria: A Survival Analysis Approach

Adedayo F. Adedotun[1*], Oluwaseun A. Odusanya[2], Hilary I. Okagbue[1], Opeyemi P. Ogundile[1]

[1] Department of Mathematics, Covenant University, Ota 112101, Nigeria
[2] Department of Mathematics, D.S Adegbenro (ICT) Polytechnic, Itori, Ogun 112106, Nigeria

Corresponding Author Email: adedayo.adedotun@covenantuniversity.edu.ng

ABSTRACT

The goal of this study is to look at the survival time distribution for diabetes patients at the National Hospital Abuja, taking into account a variety of variables. The Kaplan Meier-estimator indicated that there is no statistically significant difference in the distribution of survival time by sex, despite the fact that married patients were seen to live longer than single patients. Patients in urban and rural areas had the same estimated survival distribution after testing. It is observed that the Cox proportional model was significant when tested since the p-value = 0.000 was less than the 0.05 threshold. The distribution of survival time for patients with diabetes is shown to be substantially different for patients of the four age categories included in the study, indicating that the relative risk of patients is based on age. Every patient is predicted to acquire the danger at about the same time, with no sex-related multiplication impact. It was found out that the disease's prevalence is unaffected by several of the variables studied, indicating that more regular medical checks are required.

1. INTRODUCTION

Time to event data is often the subject of survival analysis [1]. It includes of procedures for positive-valued random variables, such as time to death, time to commencement (or relapse) of an illness, and length of stay in a hospital, in the broadest sense [2, 3]. The term "survival time" denotes the amount of time (t) that passes between a precise start time and the occurrence of a specific event or end-point [4]. In biomedical research, an event could be anything from death to illness remission to the occurrence of an epileptic episode.

Assessing the association between survival time and some biological, socio-economic, and demographic parameters that could conceivably affect the survival status of patients has become a common element of survival time data analysis, especially in medical research. Standard linear regression methods are not feasible in predicting such a connection due to censoring [5]. When the exact time a subject experienced the event of interest is unknown, censorship happens [6]. The Kaplan Meier estimator and the Cox regression model are two popular formulations [7].

The Kaplan Meier estimator is based on individual survival times and assumes that censoring is unrelated to the cause of failure [8]. The Kaplan–Meier Estimator (K-M) is a statistical method for analyzing survival data [9]. It is used to look at the distribution of patient survival times after they've been enrolled in the study. The proportion of patients still alive at a given time after recruitment or admission into the research is expressed in the analysis. Non parametric maximum likelihood estimator is another name for the K-M estimator. It's used to figure out how likely you are to live. The approach calculates the chance of dying at a specific moment in time based on survival, up until the patient is censored. As a result, it makes the most of the available data on the research sample's time to incident.

The Cox (1972) proportional hazards model is a prominent regression model that is frequently used in the area of survival research. The $h(t)$ which is the hazard function is defined as the likelihood of suffering a failure event in the infinitesimally short interval $(t, t+\Delta t)$, given that no such event has occurred previous to t.

The Cox proportional hazard model is a gold standard method for analyzing survival and event history data. Cox regression (also known as proportional hazards regression) is a technique for determining the impact of various variables on the time it takes for an event to happen.

This is known as Cox regression for survival analysis in the context of an outcome such as death. Although the method does not presuppose a specific "survival model," it is not fully nonparametric because it does assume that the effects of the predictor variables on survival are stable across time and additive on one scale [2, 8].

However, its functional form for the covariate dependence of survival time is totally parametric [10]. The regressors are, in other words, linearly connected to the log hazard. Most of the time, when modeling survival time data, the true hazard is unknown or complex, in which case the assumptions of a parametric model may not hold true. Even when all of the parametric assumptions are met, the Cox model could be as efficient as parametric models such as the Weibull model with proportional hazard [11].

Diabetes is widely acknowledged as an increasing epidemic that affects nearly every country, age group, and economy on the planet [12]. Diabetes complications are common in both type 1 and type 2 diabetes patients, but they also cause major morbidity and mortality [13]. Diabetes is a chronic condition for which there is no known cure unless in very particular circumstances, according to the literature [14]. The goal of this

disease's treatment is to keep blood sugar levels as close to normal as possible without producing low blood sugar [15]. A healthy diet, exercise, weight loss, and the use of appropriate medications (insulin in type 1 diabetes, oral drugs and perhaps insulin in type 2 diabetes) may typically achieve this [16, 17].

However, in Nigeria, the majority of studies have focused on disease prevention and variables that raise the risk of developing the disease. Because there has been little research on diabetic patient survival time, this study used the Cox proportional hazard model to investigate diabetic patient survival time. The log rank test will be used to compare the variable categories. Kaplan-Mier estimators were used to predict diabetic patient survival curves.

This study aims to use survival analysis to assess the reported case of diabetes in Nigeria with some covariates and Kaplan Meier estimator to examine the distribution of survival time for diabetes patients with their covariates and also fit a cox proportional hazard model.

This will subsequently be followed by literature review, materials and methods, results and finally, conclusion.

2. LITERATURE REVIEW

Adeloye et al. [18] carried a study to estimate country-wide and zonal prevalence, hospitalization, and mortality rates of T2DM in Nigeria. They found an increasing burden of diabetes with many people undiagnosed.

With the help of Global Moran's I, Local Moran's I, and spatial regression approaches, Olorunfemi and Ojewole [3] explored the geographical differences in Diabetes Mellitus (DM) prevalence in Nigeria. The state of Enugu has the highest prevalence of diabetes in the country. Not only was there evidence of DM clustering (I =0.30, z=3.49; p 0.05), but there was also evidence of the existence of a DM pocket in Nigeria's southeastern region, which included the states of Abia, Anambra, Enugu, and Imo. Obesity and educational attainment together account for 31 percent (OLS model) and 33 percent (Spatial error model) of the variation in the regional distribution of diabetes in the United States.

Sabir et al. [11] carried out a study to determine the prevalence of DM and its correlates in the suburban population of Northwest Nigeria. Two hundred and eighty participants were recruited using a multistage sampling technique [19-21].

Majority of studies have focused on disease prevention and variables that raise the risk of developing the disease.

3. MATERIALS AND METHODS

The survival times of 453 diabetic patients admitted to the National Hospital Abuja are examined. The failure time is defined as the time from diagnosis to death, while those whose records read "living" were right-censored because they had not died at the time of the study. Approximately 45 percent of the patients were censored, which, on average, indicates that hospital admissions of patients might end in diabetes-related mortality.

Several variables were taken into account, including sex, age, marital status, blood sugar level, diabetes type, and geographic location at the time of diagnosis. The log-rank statistic, a Chi-square type statistic, is also used to compare survival rates between male and female patients, as well as across age groups.

3.1 Cox regression

Cox regression is also used in the analysis. The hazard function, indicated by $h(t)$, represents the Cox model. In a nutshell, the hazard function is the probability of dying at the time of death. Estimated below as:

$$h(t) = h_0(t) \times exp(b_1 x_1 + b_2 x_2 + \ldots + b_p x_p) \qquad (1)$$

3.2 Kaplan Meier estimator

Assume that t_j, $j=1,2,...,n$ represents the complete set of failure times recorded (with t^+ being the highest failure time), d_j represents the number of failures at time t_j, and r_j represents the number of people at risk at time t_j. The number of people present at the start of each time period is modified based on the number of people who were censored and the number of people who experienced the event of interest in the preceding time period and failures are considered to occur first in ties between failures and censored observations. The survival function is given as:

$$\hat{S}(t) = \prod_{j:t_j \le t} \left(\frac{r_j - d_j}{r_j} \right) \; for \; 0 \le t \le t^+ \qquad (2)$$

3.3 Cox proportional model

$$\lambda(t_i, Z_i) = \lambda_0(t) exp(\beta' Z_i) \qquad (3)$$

$$\frac{\lambda(t_{(i)}, Z_{(i)})}{\sum_{j \in R(t_{(i)})} \lambda(t_{(i)}, Z_{(i)})} \qquad (4)$$

which from Eq. (3), is

$$\frac{exp(\beta' Z_{(i)})}{\sum_{j \in R(t_{(i)})} exp(\beta' Z_{(i)})} \qquad (5)$$

On the assumption of no tied events, gave the partial likelihood function as

$$L(\beta) = \prod_{i=1}^{r} \left[\frac{exp(\beta' Z_{(i)})}{\sum_{j \in R(t_{(i)})} exp(\beta' Z_{(i)})} \right] \qquad (6)$$

and the log partial likelihood is

$$\log L(\beta) = \sum_{i=1}^{r} \left[\beta' Z_{(i)} - log \sum_{j \in R(t_{(i)})} exp(\beta' Z_j) \right] \qquad (7)$$

Often, ties occur in continuous survival data that are collected in days, weeks, and months. When there are only a few ties, an approximation is provided to Eq. (5) as

$$L(\beta) = \prod_{i=1}^{r} \left[\frac{\prod_{j \in R(t_{(i)})} exp(\beta' Z_j)}{\left(\sum_{j \in R(t_{(i)})} exp(\beta' Z_j) \right)^{d_i}} \right] \qquad (8)$$

The log likelihood of Eq. (7) is

$$logL(\beta) = \sum_{i=1}^{r} \left[\sum_{j \in R(t_{(i)})} \beta' Z_j - d_i log \sum_{j \in R(t_{(i)})} exp(\beta' Z_j) \right] \qquad (9)$$

A likelihood is derived, that is an improvement over Eq. (7) is given as

$$L(\beta) = \prod_{i=1}^{r}\left[\frac{\prod_{j\in R(t_{(i)})}\exp(\beta'Z_j)}{\prod_{i=1}^{r}\left(\sum_{j\in R(t_{(i)})}(\beta'Z_j) - \frac{k=1}{d_i}\sum_{j\in R(t_{(i)})}(\beta'Z_j)\right)}\right] \quad (10)$$

with log likelihood

$$\log L(\beta) = \sum_{i=1}^{r}\left[\sum_{j\in R(t_{(i)})}\beta'Z_j - \right.$$
$$\left.\sum_{k=1}^{d_i}\log\sum_{j\in R(t_{(i)})}(\beta'Z_j) - \frac{k=1}{d_i}\sum_{j\in R(t_{(i)})}(\beta'Z_j)\right] \quad (11)$$

The maximum partial likelihood estimators $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ can be obtained for Eq. (7) and Eq. (11) from the solution of the estimating equation involving the score statistics.

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta_k} = 0, k = 1, \dots, p \quad (12)$$

and the information matrix can be obtained from

$$I(\beta) = \frac{\partial^2 \log L(\beta)}{\partial \beta_k \partial \beta_h} \quad (13)$$

Using Eq. (12) and Eq. (13), $\hat{\beta}$ can be obtained by solving the iterative equation:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^m + I^{-1}(\hat{\beta}^m)U(\hat{\beta}^m) \quad (14)$$

The variables, on the other hand, affect multiplicatively on the hazard rate thanks to the exponential link function. The effect of time-constant covariates implies that the hazard rates for any two people are proportionate, which is why the Cox model is referred to as a proportional hazards model. Suppose that Z_i and Z_j denote the covariate vectors of two individuals i and j, then the ratio of the hazard rates of these individuals is given by

$$\frac{\lambda(t_i, Z_j)}{\lambda(t_j, Z_j)} = \exp(\beta'Z_i - \beta'Z_j) \quad (15)$$

$$\lambda(t|Z_i) = c.\lambda(t|Z_j) \quad (16)$$

where, $c = \exp(\beta'Z_i - \beta'Z_j)$.

4. RESULTS

Table 1 reveals that female patients had a somewhat longer life time than male patients, with a mean survival time of 5.080 years against 5.915 years for their male counterpart. The same pattern can be seen in median survival time, with females living for an average of 5.401 years compared to males who live for an average of 4.611 years. There is no significant difference between male and female patients, according to the log-rank test statistic $\chi^2 = 0.879$ (p=0.436).

Table 1. Male and female diabetic patients' survival time means and medians

Gender, Male=1, Female=0	Mean Estimate	Median Estimate	Log Rank Mantel-COX	p-value =0.436
			0.879	
0	5.080	5.401		
1	5.915	4.611		
Overall	4.061	5.111		

Table 2. Male and female patients with diabetes: Means and medians of survival time

Marital, Married=1, single=0	Mean Estimate	Median Estimate	Log Rank Mantel-COX	p-value=0.000
			31.814	
0	4.196	4.311		
1	5.251	5.321		
Overall	4.061	5.111		

Table 2 shows that married patients live longer than single patients (reference category), with a mean survival time of 4.196 years vs 5.251 years for the reference group. The same pattern can be seen in median survival time (married patients have a median survival time of 5.321 years compared to single patients who have a median survival time of 4.311). A significant difference between male and female patients is indicated by the log-rank test result $\chi^2 = 31.814$ ($p = 0.000$).

Patients in rural regions live longer than those in urban areas (reference group), with a mean survival time of 5.125 years compared to 4.992 years for those in urban areas (Table 3). The same pattern may be seen in median survival time (rural patients live longer than urban patients, with a median of 5.211 years against 4.611 years). There is no significant difference between male and female patients, according to the log-rank test result $\chi^2 = 0.583$ ($p = 0.503$).

The Omnibus Test of Model Coefficients in Table 4 explains the 'goodness of fit' test and indicates the wellness of the model and compares it with the previous model. The significant value is less than .05 (.000) which indicates the support to the fitness of the model. In the analysis, chi-square value is 89.46 at significance level of .000 with 7 degrees of freedom.

Table 3. Survival time means and medians for diabetic patients in urban and rural areas

Area, urban=1, rural = 0	Meana Estimate	Median Estimate	Log Rank Mantel-COX	P-value= 0.503
			0.583	
0	5.125	5.211		
1	4.992	4.611		
Overall	4.061	5.111		

Table 4. Cox Proportional Hazards Model Summary

	Omnibus Tests of Model Coefficients								
-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
3592.874	89.460	7	.000	87.539	7	.000	87.539	7	.000

Table 5. Table of coefficients and hazard rates

Variables in the Equation						
	B	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**
Sex	.461	.240	7.406	1	.022	**2.520**
Marital	-.256	.291	.763	1	.520	**.976**
Area	-.161	.241	.250	1	.811	**.862**
BLS	.269	.107	3.701	1	.212	**3.282**
Age4	3.375	.444	57.388	1	.001	**8.734**
Age3	2.784	.418	30.888	1	.020	**6.439**
Age2	.985	.301	10.154	1	.033	**3.487**
Age1			.	0	.	

Table 5 shows that the estimated relative risks for male vs female patients are 2.520, implying that male patients have a 2.520 times higher chance of dying from diabetes than female patients. For married vs single patients is 0.976, which means that married patients have a 0.976 times higher chance of dying from diabetes than single patients. For example, for every 100 married patients, there are 98 more single patients who die from diabetes. The risk of dying from diabetes in urban patients is 0.862 times that of rural patients, implying that the risk of dying from diabetes in urban patients is 0.862 times that of rural patients.

The risk of dying from diabetes is 3.487, 6.439, and 8.734 for individuals aged 23-39 years (Age 2), 40-55 years (Age 3), and >55 years (Age 4), respectively, compared to the baseline age group (23 years) (Age 1). This means that older individuals have a higher chance of dying from diabetes than younger people.

5. CONCLUSION

This study discovered variables/factors that are significantly linked to a higher chance of death. Identifying patients at a higher risk of death provides the advantage of ensuring that the risk group receives special attention during their follow-up to reduce the risk of mortality while on therapy.

When comparing the time of hospital admission to death, based on diabetes data, male diabetic patients are seen to have an approximately equal survival time with their female counterparts, married diabetic patients survive diabetes longer than single patients, which could be attributed to the better diet control that married diabetic patients have over single diabetic patients. Although the survival distribution estimates for patients in urban and rural regions are similar, we do notice that individuals in rural areas live a bit longer than those in urban areas. Furthermore, a relative risk (hazard ratio) of male vs female (baseline) of larger than one (1) indicates that male patients have a higher risk of dying from diabetes than female patients, indicating that female patients have a better chance of surviving. The chance of dying from diabetes increases with age for each age group.

The study find support from the work of ref. [22], who opined that shorter survival time in female patients than male and no gender difference in survival of diabetes.

There are several limitations to consider in this study. First, all factors are self-reported, thus possibility subject to measurement error. Due to the complex nature of the modeling process, the models were kept relatively simple to ensure convergence

REFERENCES

[1] Leger, S., Zwanenburg, A., Pilz, K., Lohaus, F., Linge, A., Zöphel, K., Richter, C. (2017). A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. Scientific Reports, 7(1): 1-11. https://doi.org/10.1038/s41598-017-13448-3

[2] Odetunmibi, O.A., Adejumo, A.O., Anake, T.A. (2019). Log-linear modelling of effect of age and gender on the spread of hepatitis b virus infection in Lagos state, Nigeria. Open Access Macedonian Journal of Medical Sciences, 7(13): 2204. https://doi.org/10.3889/oamjms.2019.573

[3] Olorunfemi, O., Ojewole, F. (2019). Medication belief as correlate of medication adherence among patients with diabetes in Edo State, Nigeria. Nursing Open, 6(1): 197-202. https://doi.org/10.1002/nop2.199

[4] Manoharan, T. (2017). Inferential Procedures on Log-normal Model for Left-Truncated and case-k Interval Censored Data with Covariates.

[5] Rossello, X., González-Del-Hoyo, M. (2021). Survival analyses in cardiovascular research, part I: the essentials. Revista Española de Cardiología (English Edition), 75(1): 67-76. https://doi.org/10.1016/j.rec.2021.06.003

[6] Hazra, A., Gogtay, N. (2017). Biostatistics series module 9: Survival analysis. Indian Journal of Dermatology, 62(3): 251-257. https://doi.org/10.4103/ijd.IJD_201_17

[7] McGregor, D.E., Palarea-Albaladejo, J., Dall, P.M., Hron, K., Chastin, S.F.M. (2020). Cox regression survival analysis with compositional covariates: application to modelling mortality risk from 24-h physical activity patterns. Statistical Methods in Medical Research, 29(5): 1447-1465. https://doi.org/10.1177/0962280219864125

[8] Papatheodorou, K., Banach, M., Bekiari, E., Rizzo, M., Edmonds, M. (2018). Complications of diabetes 2017. Journal of Diabetes Research, 2018: 3086167. https://doi.org/10.1155/2018/3086167

[9] Etikan, İ., Abubakar, S., Alkassim, R. (2017). The Kaplan-Meier estimate in survival analysis. Biom Biostatistics Int J, 5(2): 00128.

[10] Osayomi, T. (2019). The emergence of a diabetes pocket in Nigeria: The result of a spatial analysis. GeoJournal, 84(5): 1149-1164. https://doi.org/10.1007/s10708-018-9911-2

[11] Sabir, A.A., Balarabe, S., Sani, A.A., Isezuo, S.A., Bello, K.S., Jimoh, A.O., Iwuala, S.O. (2017). Prevalence of diabetes mellitus and its risk factors among the suburban population of Northwest Nigeria. Sahel Medical Journal, 20(4): 168. https://doi.org/10.4103/smj.smj_47_16

[12] Omaku, P.E., Adamu, N.T.Y.I. Survival Analysis of Reported Cases of Diabetes Disease in Nigeria.

[13] Holman, N., Knighton, P., Kar, P., O'Keefe, J., Curley, M., Weaver, A., Valabhji, J. (2020). Risk factors for COVID-19-related mortality in people with type 1 and type 2 diabetes in England: A population-based cohort

study. The lancet Diabetes & Endocrinology, 8(10): 823-833. https://doi.org/10.1016/S2213-8587(20)30271-0

[14] Davis, J., Penha, J., Mbowe, O., Taira, D.A. (2017). Peer reviewed: Prevalence of single and multiple leading causes of death by race/ethnicity among US adults aged 60 to 79 years. Preventing Chronic Disease, 14: E101. https://doi.org/10.5888/pcd14.160241

[15] Jangid, H., Chaturvedi, S., Khinchi, M.P. (2017). An overview on diabetis mellitus. Asian Journal of Pharmaceutical Research and Development, 1-11. http://www.ajprd.com/index.php/journal/article/view/30 2.

[16] Awuchi, C.G., Echeta, C.K., Igwe, V.S. (2020). Diabetes and the nutrition and diets for its prevention and treatment: A systematic review and dietetic perspective. Health Sciences Research, 6(1): 5-19.

[17] Okagbue H.I., P.E. Oguntunde P.E., Bishop S.A., Obasi E.C.M., Opanuga A.A., Ogundile, (2020). Review on the reliability of medical O.P. Contents on YouTube. International Journal of Online and Biomedical Engineering, 16(1): 83-99. https://doi.org/10.3991/ijoe.v16i01.11558

[18] Adeloye, D., Ige, J.O., Aderemi, A.V., Adeleye, N., Amoo, E.O., Auta, A., Oni, G. (2017). Estimating the prevalence, hospitalisation and mortality from type 2 diabetes mellitus in Nigeria: A systematic review and meta-analysis. BMJ Open, 7(5): e015424. http://dx.doi.org/10.1136/bmjopen-2016-015424

[19] Ran, L., Hu, T., Gao, S. (2020). Estimation of covariate effects in proportional cross-ratio model of bivariate time-to-event outcomes. Communications in Statistics-Simulation and Computation, pp. 1-15. https://doi.org/10.1080/03610918.2020.1839093

[20] Wang, P., Li, Y., Reddy, C.K. (2019). Machine learning for survival analysis: A survey. ACM Computing Surveys (CSUR), 51(6): 1-36. https://doi.org/10.1145/3214306

[21] Wulandari, I., Kurnia, A., Sadik, K. (2021). Weibull regression and stratified cox regression in modelling exclusive breastfeeding duration. In Journal of Physics: Conference Series, 1940(1): 012001. https://doi.org/10.1088/1742-6596/1940/1/012001

[22] Akindutire, O.R., Ogunlade, T. (2020). Risk factor identification using survival analysis. International Journal of Mathematics Trends and Technology (IJMTT), 66(3): 104-109. https://doi.org/10.14445/22315373/IJMTT-V66I3P516