Risk Factor Identification Using Survival Analysis

Akindutire O. R[1] and Ogunlade T. O[2]

1. *Department of Acturial Science, Ekiti State University, Ado-Ekiti, Nigeria*
2. *Department of Mathematics, Ekiti State University, Ado-Ekiti, Nigeria*

Abstract*: In the 1960s, Diabetes was considered insignificant in Nigeria. However, its increase prevalence has been contributed to by major socio-economic changes, with 12% cases recorded in some areas. The study assesses the contribution of some risk factors to the survival diabetes patients. Cox proportional hazards regression model, Long-rank and Kaplan- Meier methods were employed in analyzing data collected from a tertiary hospital. The Kaplan Meier curves indicate consistently better survival prognosis in Type 1 diabetes patients than type 2 and gestational diabetes. The long-rank test compared survival function of various levels of the covariates. It was observed that survival experience of different gender is not significantly different while that of various years of admission and diabetes categories did at 0.05 level of significance. Hazard ratio also revealed short survival time in patients diagnosed with type 1 and 2 diabetes relative to baseline hazard (gestational diabetes). However, diagnosis does not contribute significantly to the model. Based on the data, it has been established to a usable extent that better survival prognosis in type 1 diabetes patients than those with type 2 or gestational diabetes, Shorter survival time in female patients than male and no gender difference in survival of diabetes. The only significant factor amongst gender, year of admission and diagnosis was year of admission and earlier admission may be free of event.*

Keywords*: Survival analysis, diabetes, hazard function, regression coefficients.*

I. INTRODUCTION

Diabetes Mellitus (DM) is a serious threat to global health, a major cause of morbidity and mortality in countries of all income levels [1-3]. DM is a chronic non-communicable disease (NCD) identified with high levels of blood sugar or glucose, leading gradually to serious damage to the heart, blood vessels, eyes, kidneys, and nerves [1]. DM occurs when the pancreas produces insufficient insulin[1] or when the body cannot effectively use the insulin produced [4]. DM has 3 major types; the type 1 diabetes (insufficient insulin production), type 2 diabetes (ineffective use of insulin) and gestational diabetes [1]. DM is a major cause of heart attacks, stroke, kidney failure, blindness and lower limb amputation [4]. Most common type of DM in adults is the type 2 diabetes [1]. Adult with diabetes are 2 or 3 times more likely to develop heart attack and strokes [2]; 3 times at risk to tuberculosis [5]; DM is associated with substantial premature death from several cancers [6].

International Diabetes Federation (IDF) estimates current prevalence of diabetes (type 1 and type 2 combined, both diagnosed and undiagnosed) in people aged 20–79 years to be 463 million (9.3% of persons in this age group) against 151 million (4.6% of persons in this age group) in 2000 [3]. In 2014, global prevalence of DM among adults (over 18 years of age) was 8.5% (422 million) compared to 4.7% (108 million) in 1980, a 3.8% increase [2, 4]. It is projected that 578 million people will have diabetes by 2030 if the pandemic is not addressed [3]. Middle- and low-income countries have experienced higher increasing prevalence of DM. DM was estimated to be the seventh leading cause of death, accounting for an estimated 1.6 million deaths worldwide in 2016 [4]. DM attributes for 2.6% of global blindness through diabetic retinopathy [7]. 10% of global health expenditures is spent on diabetes [3]. Diabetes prevalence is higher in urban areas than the rural areas [3, 8, 9]. IDF reports North America and Caribbean with the highest raw diabetes prevalence (13.3%) and highest age-adjusted comparative diabetes prevalence in Middle east and north Africa (12.2%) [3].

In sub-Saharan Africa (SSA), prevalence and burden of type 2 diabetes (T2DM) are severe and rapidly increasing [8, 10], accounting for over 90% of diabetes in the region [8]. Several factors such as; lack of funding for NCDs, dearth of studies and recommendations specific to the African population, large disparity in socio-economic factors and inequity in health care (public and private) constitute distinct challenges in combating diabetes [9]. DM has the highest impact in developing countries, also mortality and morbidity related to the disease is highest in SSA than any other region [9, 11]. A 2011 study reported a very high (>10%) prevalence rate of T2DM in Zimbabwe, 12% in urban Kenya, low to medium prevalence rate (0-7%) in Cameroon, Ghana, Guinea, Kenya, Nigeria, South Africa and Uganda; with a low 0.6% in rural Uganda [8]. IDF reports raw diabetes prevalence (%) and age-adjusted comparative diabetes prevalence in SSA to be 3.9 and 4.7% respectively [3].

In 1992, a nationwide population survey of DM was carried out by the Federal ministry of Health, with a national prevalence estimated at 2.2% [12]. Since then, there has been no national health survey on diabetes [13]. Various study on prevalence of DM in Nigeria ranged from a low level (0.8%) in Ibadan to middle (7-8%) in Port Harcourt and Maiduguri and a high (12%) among adults [13-16]. In the 1960s, diabetes was considered rare in Nigeria [17], socio-economic changes might have contributed to increased prevalence. IDF estimates 2.7 million people (aged 20-79 years) with diabetes in Nigeria[3]. Arguments about gross "under-report" has been made on IDF's estimate due to the estimates been derived from the extrapolation of data from other countries [13]. However, determining the survival rate of DM becomes imperative to know the mortality burden and also facilitate apt allocation and planning.

II. METHODS

The data are secondary data *obtained from the records of five hundred and seventy six (576) registered diabetes patients at University of Ilorin Teaching Hospital, Ilorin Nigeria.* Data collected is based on the length of stay, age, gender, type of diabetes (Type 1, Type 2 and Gestational Diabetes) for a four (4) year period i.e. 2010-2013 respectively.

Table 1.0: Data Presentation

Variables	Covariates
Types of Diabetes	Type1, Type2, and Gestational Diabetes
Gender of Diabetes Patients	Male and Female
Status of Diabetes Patients	Dead and Alive
Year of Admission	2010,2011,2012 and 2013

Summary of Data Presentation

Table 2.0: Frequency table showing gender of patients

GENDER	**Frequency**	**Events**	**Censored**		**Valid Percentage**	**Cumulative Percentage**
			N	**Percentage**		
Male	276	96	180	65.2%	47.9%	47.9%
Female	300	104	196	65.3%	52.1%	100.0%
Total	576	200	376	65.2%	100.0%	

Table 2.1: Frequency table showing Diabetes category of patients

DIABETES CATEGORY	Frequency	Events	Censored		Valid Percentage	Cumulative Percentage
			N	Percentage		
Gestational	79	22	57	72.2%	13.7%	13.7%
Type 1	266	65	201	75.6%	46.2%	59.9%
Type 2	231	113	118	51.1%	40.1%	100.0%
Total	576	200	376	65.3%	100.0%	

Table 2.2: Frequency table showing year of admission of patients

YEAR	Frequency	Events	Censored		Valid Percentage	Cumulative Percentage
			N	Percentage		
2010	101	85	16	15.8%	17.5	17.5
2011	161	34	127	78.9%	28.0	45.5
2012	279	67	212	76.0%	48.4	93.9
2013	35	14	21	60.0%	6.1	100.0
Total	576	200	376	65.3%	100.0%	

The two functions that were used to summarize, describe and analyze survival data were the survival function and hazard function. Survival function indicates the probability that the event of interest has not yet occurred by time t; thus, if T denotes time until death, S(t) denotes probability of surviving beyond time t. In practice, if there are no censored observations, the survivorship functioning estimated as the proportion of patients surviving longer than *t:*

$$\hat{S}(t) = \frac{\text{number of patients surviving longer than t}}{\text{total number of patients}} \quad \ldots\ldots\ldots\ldots \quad (1)$$

The survivor function is often expressed as a Kaplan-Meier curve. When there is a vertical drop in a Kaplan-Meier curve, this indicates an event. Thus, the survival function is given by:

$$\mathrm{S(t)} = \Pr = 1 - \mathrm{F(t)} = \int_t^{\infty} f(x)dx \quad \ldots\ldots\ldots\ldots \quad (2)$$

The hazard function, conventionally denoted λ, is defined as the event rate at time t conditional on survival until time t or later (that is, $T \geq t$). Suppose that an item has survived for a time t and we desire the probability that it will not survive for an additional time *dt*. It is a measure of instantaneous potential whereas a survival curve is a cumulative measure over time; it may be used to identify a specific model form, such as an exponential, a Weibull, or a lognormal curve that fits one's data. The hazard function can also be defined in terms of the cumulative distribution function $F(t)$ and the probability density function $f(t)$:

$$h(t) = \frac{\mathrm{f(t)}}{1-\mathrm{F(t)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

III. RESULTS

Fig. 1: Results of the Kaplan-Meier graphs for the categorical variables;

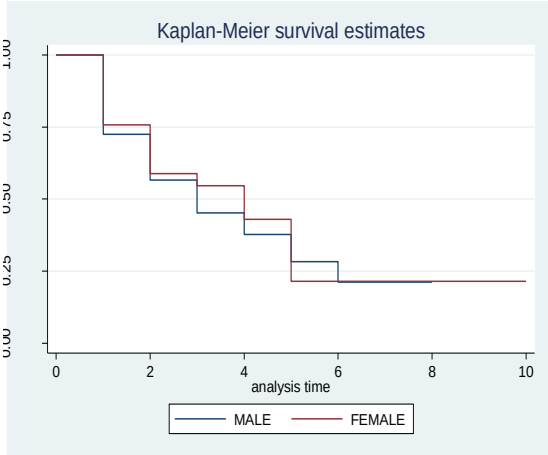

Fig 1.1: Survival curve based on gender of diabetes patients

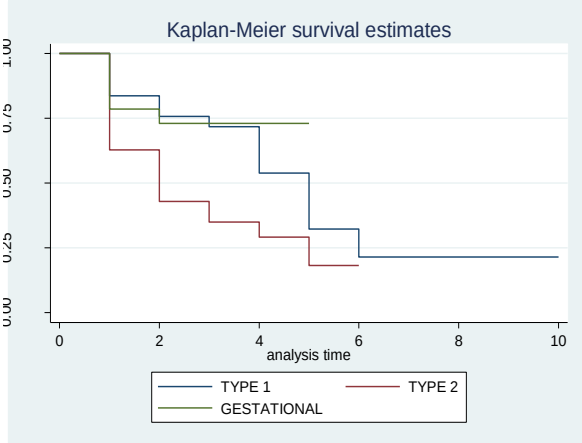

Fig 1.2: Survival curve based on type of diabetes disease

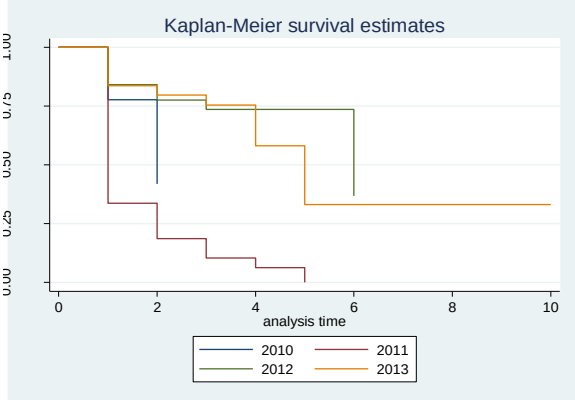

Fig 1.3: Survival curve based on year admitted

Figure 1.1 shows the K-M curves for gender of two groups which appear to be different. Also, figure 1.2 indicates the KM curves for the categories of diabetes in which type 1 diabetes have a close relation better survival prognosis (i.e., maintenance) than patients with type 2 and gestational diabetes. The other two groups are not very different from each other, although type 2 diabetes has a better prognosis than gestational diabetes. While figure 1.3 revealed that K-M curves for the four years which are quite different.

Table 3: Results of the log-rank test for the categorical variables of diabetes patients:

Covariates	d.f	Chi-square	p-value
GENDER	1	0.71	0.399
DIAGNOSIS	2	35.03	0.201
YEAR	3	131.44	0.001

The Cox's Proportional Hazard Model was employed to determine the hazard ratio of the various covariates. The results indicate that only significant factor out of gender, year of admission and diagnosis was year of admission which was associated with the survival of diabetes and the result shows that those who were admitted at an earlier time were less likely to get the event. The result obtained by the PHreg procedure is shown in the table below:

Table 4. Summary statistics for the PHR model for the categorical variables of diabetes patients:

Covariates	Hazard Ratio	Coefficients (β)	**Std. Err**	**d.f**	**Z**	**P>(z)**	**95% confident Interval**	
Female	0.825227	-0.1920	0.1191415	1	-1.33	0.183	0.6218447	1.095128
Type 1 diabetes	0.7979588	0.1294	0.215426	2	-0.84	0.403	0.4700894	1.354505
Type 2 diabetes	1.139049	-0.2213	0.2471621	2	0.60	0.549	0.7444581	1.742788
2011	0.4942281	0.6596	0.1201263	3	-2.90	0.004	0.3069274	0.795828
2012	0.4427774	-0.8147	0.1090504	3	-3.31	0.001	0.2732407	0.7175061
2013	1.931786	-0.7048	0.409604	3	3.11	0.002	1.274898	2.927133

Model:

$h_o(\mathrm{t})e^{(-0.1920 female + 0.1294 type\ 1\ diabetes - 0.2213 type\ 2\ diabetes + 0.6596 year\ 2 - 0.8147\ year\ 3 - 0.7048\ year\ 4)}$

The estimated regression coefficients, (coef), and the hazard ratio (exp (coef)) between the groups of covariates were obtained. Therefore, the hazard ratio which we will use to interpret the cox proportional hazards model is compared based on its closeness to 1.

Thus, the hazard ratio for male relative to female is 0.825. Since this ratio is less than 1, it implies that male patients have a shorter survival time than the female patients and the risk of dying from diabetes by male patients is 0.825 times that of female patients. It was observed that the risk of dying based on gender is not significant. The hazard ratio for different levels of the year admitted for treatment relative to the reference hazard (year 2010) are 0.494, 0.443, 1.932 for years 2011, 2012, and 2013 respectively. This shows that patients admitted for treatment in year 2011, 2012 and 2013 have a lower risk and survive longer than the baseline hazard. However, it was observed that year contribute significantly to the model.

In conclusion, the hazard ratio for different diabetes categories were obtained, which revealed that patients diagnosed with Type 1 diabetes and Type 2 diabetes have a short survival time (i.e. will die faster from diabetes) relative to the baseline hazard (Gestational diabetes). However, it was observed that diagnosis does not contribute significantly to the model.

REFERENCES

[1] World Health Organization. *Diabetes*. 2019 [cited 2020 02/01/2020]; Health-topics Diabetes]. Available from: https://www.who.int/health-topics/diabetes.

[2] Sarwar, N., Gao, P., Seshasai, S. R., Gobin, R., Kaptoge, S., Di Angelantonio, E., ... & Stehouwer, C. D. (2010). Emerging Risk Factors Collaboration Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*, *375*(9733), 2215-2222.

[3] International Diabetes Federation, *IDF Diabetes Atlas Ninth edition 2019*, in *IDF Diabetes Atlas*. 2019, International Diabetes Federation: Brussels, Belgium. p. 176.

[4] World Health Organization. *Diabetes Fact Sheet*. 2018 30/10/2018 [cited 2020; Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes.

[5] Jeon, C.Y. and M.B. Murray, *Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies*. PLoS medicine, 2008. **5**(7): p. e152.

[6] Emerging Risk Factors Collaboration. (2011). Diabetes mellitus, fasting glucose, and risk of cause-specific death. *New England Journal of Medicine*, *364*(9), 829-841.

[7] Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, Jonas JB, Keeffe J, Leasher J, Naidoo K, Pesudovs K, Resnikoff S, Taylor HR; Vision Loss Expert Group., *Causes of vision loss worldwide, 1990–2010: a systematic analysis*. The lancet global health, 2013. **1**(6): p. e339-e349.

[8] Hall V, Thomsen R.W, Henriksen O & Lohse N ., *Diabetes in Sub Saharan Africa 1999-2011: epidemiology and public health implications. A systematic review.* BMC public health, 2011. **11**(1): p. 564.

[9] Pastakia, S. D., Pekny, C. R., Manyara, S. M., & Fischer, L. (2017). Diabetes in sub-Saharan Africa–from policy to practice to progress: targeting the existing gaps for future care for diabetes. *Diabetes, metabolic syndrome and obesity: targets and therapy*, *10*, 247

[10] Mbanya JC, Motala AA, Sobngwi E, Assah FK, Enoru ST.., *Diabetes in sub-saharan africa.* The lancet, 2010. **375**(9733): p. 2254-2266.

[11] Shaw, J.E., R.A. Sicree, and P.Z. Zimmet, *Global estimates of the prevalence of diabetes for 2010 and 2030.* Diabetes research and clinical practice, 2010. **87**(1): p. 4-14.

[12] Akinkugbe, O. and O. Akinyanju, *Non-communicable diseases in Nigeria: final report of a national survey.* Lagos: Federal Ministry of Health—National Expert Committee on Non-Communicable Diseases, 1997: p. 1-12.
[13] Uloko AE, Musa BM, Ramalan MA, Gezawa ID, Puepet FH, Uloko AT, Borodo MM, Sada KB.., *Prevalence and risk factors for diabetes mellitus in Nigeria: a systematic review and meta-analysis.* Diabetes Therapy, 2018. **9**(3): p. 1307-1316.
[14] Dahiru, T., A.A. Aliyu, and A. Shehu, *A review of population-based studies on diabetes mellitus in Nigeria.* Sub-Saharan African Journal of Medicine, 2016. **3**(2): p. 59.
[15] Gezawa ID**,** Puepet FH**,** Mub BM **,** Uloko AE**,** Bakki B**,** Talle **MA,** Haliru I ., *Socio-demographic and anthropometric risk factors for type 2 diabetes in Maiduguri, North-Eastern Nigeria.* Sahel Medical Journal, 2015. **18**(5): p. 1.
[16] Nyenwe EA, Odia OJ, Ihekwaba AE, Ojule A, Babatunde S. *Type 2 diabetes in adult Nigerians: a study of its prevalence and risk factors in Port Harcourt, Nigeria.* Diabetes research and clinical practice, 2003. **62**(3): p. 177-185.
[17] Olatunbosun S. T, Ojo P. O, Fineberg N. S. and Bella A. F, *Prevalence of diabetes mellitus and impaired glucose tolerance in a group of urban adults in Nigeria.* Journal of the National Medical Association, 1998. **90**(5): p. 293.