Malaysian Journal of Applied Sciences 2019, **Vol 4(1)**: 73-83

eISSN 0127-9246 (Online)

Malaysian Journal of Applied Sciences

ORIGINAL ARTICLE

Analysis of Diabetes Data using Extended Cox Model with Frailty under Partial and Penalized Partial Likelihood Estimation Methods.

Alfred A. Abiodun[a] , Benjamin A. Oyejola[a] and Kazeem A. Adeleke[b]

[a]Department of Statistics, University of Ilorin, PMB 1515, Ilorin, Kwara State, Nigeria.
[b]Deparrment of Mathematics, Obafemi Awolowo University, Ile-Ife, Nigeria.

*Corresponding author: alfredabiodun1@gmail.com

Received: 04/02/2019, Accepted: 09/04/2019

Abstract

Data on Diabetes were analyzed using partial likelihood (Pl) and penalized partial likelihood (Ppl) estimation methods in non-proportional hazards model with dichotomous time-varying covariates. Gamma and Inverse Gaussian frailty distributions were used to account for patient- specific unobserved heterogeneity. Four likelihood configurations were formed from the combinations of the two estimation methods and frailty distributions. These are Partial likelihood with Gamma frailty, Partial likelihood with Inverse Gaussian frailty, Penalized partial likelihood with Gamma frailty and Penalized partial likelihood with Gamma frailty. The results revealed that age and body mass index of the patients significantly increased the risk of death from diabetes, while regular exercise had significant decreased risk of death. Penalized partial likelihood estimation method generally outperformed models with Partial likelihood under all scenarios for the data and Gamma frailty provided a better fit in accounting for unobserved heterogeneity among the diabetic patients.

Keywords: Survival time; frailty; time-dependent covariate; AIC, BIC.

Introduction

Diabetes mellitus is a serious public health problem that has implications for individuals, communities, health and human services (AIHW; 2009, Comino, 2015). The increased prevalence of diabetes and its significant impact on the use of health care services, particularly hospitals, is a concern for health planners (Comino, 2015). A number of studies have demonstrated that people with diabetes (Aro et al., 1994; Bo et al., 2004; De Berardis et al.,2012) have hospital admission rates between 2 and 6 times higher than people without diabetes (Carral et al., 2002). People with diabetes also have excessive lengths of hospital stay compared to people without diabetes (Aro et al., 1994; Carral et al., 2002).

Cox proportional hazards (PH) model is the most preferred model in investigating the effect of variables on survival time. The basic assumption of the model is that the hazard rate related to different levels of the factors be constant throughout the follow-up period (Başar, 2006). This assumption is violated in many practical situations because most of the covariates are time-varying, thereby making applicability of proportional hazards models inappropriate. Non-proportional hazards often arise if some covariate only affects survival up until sometime t or if the size of its effect changes over time. Assessing the assumption of proportional hazards is a central theme in survival analysis. It is extensively discussed in several studies and statistical

textbooks (Grambsch & Therneau,1994; Khalid, et al., 2013; Colagiuri et al., 2009; Rosentha et al.,1998); Putter et al., 2005; Ng'andu, et al., 1997).

Various forms of time-varying functions are available, including t, logt and Heaviside. One popular form of time-varying covariate is discussed in Zhou (2001), and was also highlighted in Austin (2012). Such covariates are known to change at most once, say from untreated to treated. Adeleke et al. (2015) discussed a Semi-parametric non-proportional hazards model with time-varying covariates, with application to fertility data. Correctly accounting for time-varying covariates is important because it allows one to avoid the problem of survivor-treatment or immortal-time bias (Suissa, 2007; Beyersmann et al., 2008; Austin et al., 2006.

In research involving medical data, patients often differ substantially in characteristics and this causes variations in the effects of observed covariates amongst them. Also, some relevant covariates may be unavailable to the researchers (Munda, 2012), and this makes the assumption of independent and identical distribution, which is often made regarding survival data of different patients implausible. Introduction of random effect (frailty) in most cases improves the fit of such models. Vaupel et al., (1979) introduced the concept of frailty to the biostatistical community and applied it to population mortality data.

This study aims to applying non proportional hazards model with a dichotomous one-step change point heaviside function of time-varying covariates to investigate the survival time of diabetic patients.

Materials and Methods

Data were collected on 150 diabetic patients who were diagnosed and admitted at the University of Maiduguri Teaching Hospital, Borno State. From these, 60 patients died during the duration of admission and the remaining 90 were either lost to follow-up or were still alive at the time of data collection and were therefore right-censored at the time they were last seen or time the data were collected. Survival time was defined as the time from hospital admission to death due to diabetes, recorded in weeks. The covariates thought to be associated with diabetes were also collected, and these included: age of the patient in years, sex (male=1, female=2), body mass index in kg/m2, alcohol intake (yes=1, no=2), salt intake (normal=1, abnormal=2), family history (inherited=1, uninherited=2) and exercise (yes=1, no=2). Individual patient identification number was used as the frailty information.

The Cox Proportional Hazards Model

Consider the study involving n subjects, and for subject i ,i = 1,…, n, we let and Ci be the survival and censoring times respectively, be a vector of covariates, where Ti and Ci are independent given . The observed data are =min(,Ci), with censoring indicator . The proportional hazards (PH) model Cox (1972) for examining the relationship of hazard function with a set of covariates is given as.

$$\alpha_i(t|z_i) = \alpha_o(t)\exp(\gamma' z_i), \quad (1)$$

The Cox model formula in (1) says that the hazard at time t is the product of two quantities. The first of these is and is called the baseline hazard function, the second quantity is the exponential expression (exp) to the linear sum of , where the sum is over the p explanatory variables zi (Kleinbaum and Klein, 2012).

A tempting property of the model in (1) is that, even though the baseline hazard part of the model is unspecified, it is possible to estimate the regression coefficients in the exponential part of the model.

Extension of the Cox Proportional Hazards Model

The Cox PH model in (1) assumes that

$$\frac{\alpha(t|z}{\alpha(t|z*} = \theta$$

that is the hazard ratio comparing two individuals with covariates z and z* is constant over time. The violation of this assumption leads to the Extended Cox model which is obtained by including time-varying covariate , being the interaction on which of the 's the assumption is violated with time function given as

$$x_i(t) = z_i \times g(t) \tag{2}$$

Thus cox model with time-varying covariates can be written as

$$\alpha_i(t|z_i, x_i(t) = \alpha_0(t)\exp(\gamma' z_i + \beta_t' x_i(t)) \tag{3}$$

Selection of varies according to the state of the variables used and according to the information level of the researchers. In this study we have chosen as a dichotomous one-step change point Heaviside function as considered by Austin (2012) and defined by

, $$g(t) = \begin{cases} 1 \; if \;\; t \geq t_c \\ 0 \; if \;\; t < t_c \end{cases} \tag{4}$$

where is the point where the change that results in violation of proportionality assumption occurs. If = 0 for all t, then in (2) becomes 0 and the model in (3) reduces to the standard Cox model in (1). The interaction between and each of the covariates on which proportionality assumption is violated is then incorporated into the model.

For variables not satisfying the proportionality assumption, the power of the corresponding tests is reduced, that is, it is less likely to conclude for a significant effect when there is actually one. If the hazard ratio is increasing over time, the estimated coefficient assuming PH is overestimated at first and underestimated later on. For those variables of the model with a constant hazard ratio, the power of tests is also reduced as a consequence of an inferior fit of the model (Bellera et al.,2010)

.

Incorporation of Frailty

A further extension of the model in (3) is the incorporation of frailty due to individual patient's unobserved heterogeneity so that the model becomes

$$\alpha_i(t|z_i, x_i(t), \xi_i) = \xi_i \alpha_0(t)\exp(\gamma' z_i + \beta_t' x_i(t) \tag{5}$$

Two frailty distributions considered in the study are the one–parameter Gamma and Inverse Gaussian distributions. The density function for one-parameter Gamma with mean 1 and variance variance θ can be given as

$$h_\theta(\xi) = \frac{\xi^{\frac{1}{\theta}-1}\exp(-\xi/\theta)}{\Gamma(1/\theta)\theta^{\frac{1}{\theta}}}, \xi, \; \theta > 0 \tag{6}$$

where (.) denotes the Gamma function.

Inverse Gausian distribution has the density given by

$$h_\theta(\xi) = (2\pi\theta\xi^3)^{-12} \exp\left\{\frac{1}{20}\left(\xi - 2 + \frac{1}{\xi}\right)\right\} \tag{7}$$

Estimation of Parameters

Two estimation methods, Partial and Penalized-partial likelihoods were employed in the study. Suppose we have relevant information for an individual i (i =1, . . . , n) contained in the vector) with ti being the time to death from diabetes, where di is the censoring indicator which takes on value 1 if death occurs on patient i and value zero if the patient is censored. The vectors of time-fixed and time-varying covariates are zi and xi(t) respectively, where the value of xi(t) is as earlier described and is the frailty term for patient i with distribution as given in (6) and (7).

Partial Likelihood

Cox model utilizes partial likelihood method in which estimates for the parameter of interest for time fixed and βt for time-varying can be found by maximizing the partial likelihood. The likelihood is given as in standard survival models

$$l_{PL}(\gamma, \beta_t \mid \xi_i) = \prod_{i=1}^{n}\left[\frac{\xi_i \exp(\gamma' z_i + \beta_t' x_i(t)}{\sum_{R(t)} \xi_s \exp(\gamma' z_s + \beta_t' x_s(t)}\right]^{d_i} \tag{8}$$

where R(t) denotes summation over all individuals in the risk set.

Penalized-partial Likelihood

In penalized partial likelihood, the frailty density in the complete data likelihood is treated as a penalty term. Therefore, following (McGilchrist, 1993), the penalized partial log likelihood for the frailty model can be written as

$$l_{PPL}(\gamma, \beta_t \mid \xi_i) = \log\left[\left(\frac{\xi_i \exp\{\gamma' z_i + \beta_t' x_i(t)}{\sum_{R_t} \xi_{is} \exp\{\gamma_s' + \beta_t' x_{is}(t)} + \sum_{i=1}^{n} log(h_\theta(\xi_i)\right)^{d_i}\right] \tag{9}$$

Conditional on the frailty term , maximization of the penalized partial likelihood criterion leads to the same parameter estimates for the fixed effects time-varying effects and the frailty parameter θ as the EM algorithm (Therneau et al., 2003).

Data Analysis

Descriptive analysis of the data were carried out, including Kaplan-meier curves of the survival probabilities. Proportionality hazards (PH) assumption was checked for all the covariates using Schoenfeld Residuals as well as graphical method. The Schoenfeld Residuals was obtained as the observed minus expected values of the covariates at each time point. A p-value less than 0.05 is an indication of violation of proportionality assumption. For the graphical approach, the empirical plots of log–log survival probabilities for different individuals were obtained. For proportional hazards assumption to be upheld on a categorical covariate, one should expect that the plots be approximately parallel, otherwise violation of assumption is evidenced.

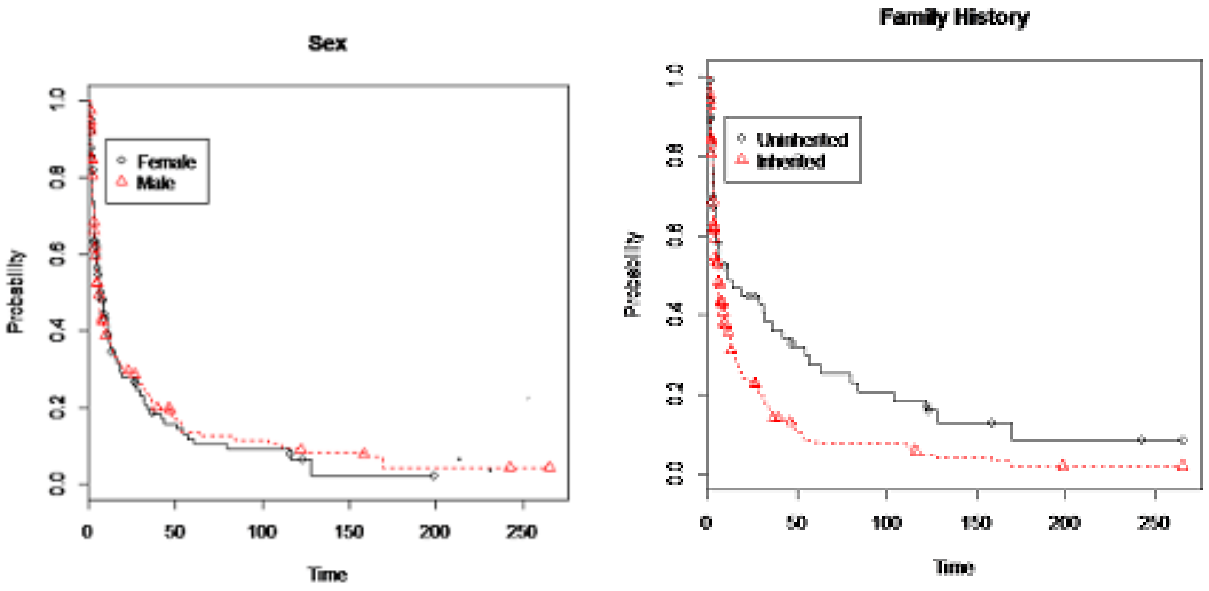
In estimating the Partial and Penalized partial likelihood parameters, the change point described in (4) was located at the median survival time which is 42 weeks in which case the time varying function was given as

$$g(t) = \begin{cases} 1 \; if \;\; t \geq 42 \\ 0 \; if \;\; t < 42 \end{cases} \tag{10}$$

The time-varying covariate was thus built with those covariates that violated proportionality assumption by their interactions with function of time) as given in (4). Non-proportional Hazards model (5) was then fitted using the four likelihood configurations, namely Partial likelihood with Gamma frailty (Pl-gam), Partial likelihood with Inverse Gaussian frailty (Pl-invgaus), Penalized partial likelihood with Gamma frailty (Ppl-gam) and Penalized partial likelihood with Inverse gaussian frailty (Ppl-invgaus).

Results and Discussion

For our dataset on diabetic patients under study, the average length of stay in the hospital was 20.0 weeks and the maximum was 266 weeks. The mean age of the patients was 43±16.6 years and the mean body mass index was 27.5±10.6 kg/m^2. Kaplan Meier curves for comparing survival probabilities are shown in Figure 1(a-e). As observed, male patients had slightly higher survival probability than females, with median survival time of 11 weeks against 6 weeks for females. Patients with uninherited family history had higher survival probability with median survival time of 11 weeks than those who did not inherit the disease, with median survival time of 7 weeks. The median survival time of those who engaged in exercise was 7 weeks against those who did not with median survival time of 6 weeks. Also, those who did not take alcohol had a median survival time of 18 weeks against 9 weeks for those who took alcohol, and those whose salt intake was normal had a higher survival probability than those who had abnormal salt intake, with median survival times of 8 weeks and 4 weeks respectively.



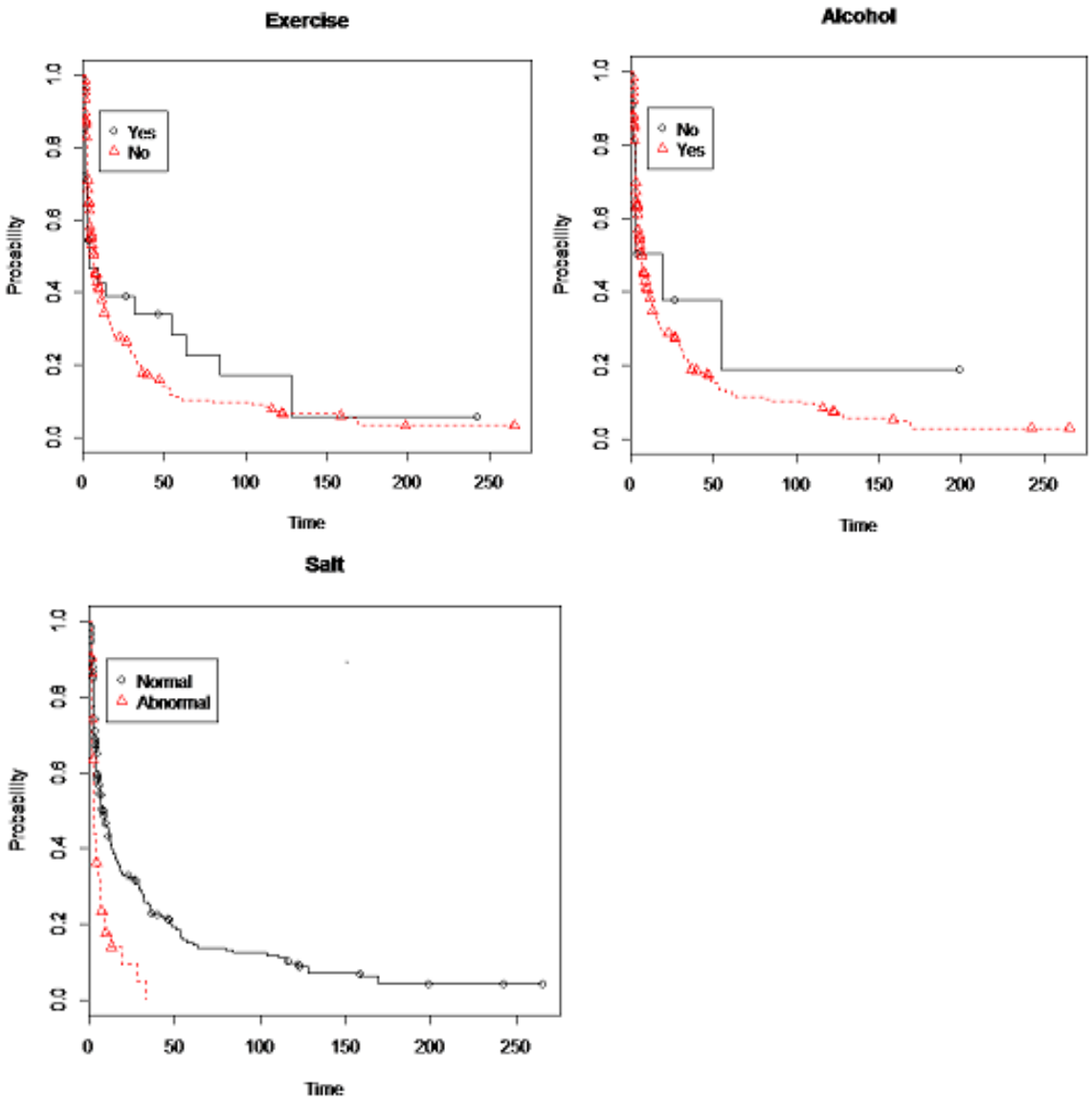


Figure 1. Kaplan Meier curves for comparing survival probabilities

Proportional Hazards Assumption check

The results of Schoenfeld Residuals are presented in Table 1. The column Rho is the Pearson product-moment correlation coefficient between the scaled Schoenfeld residuals and the function of time for each covariate in the model (both categorical and continuous). The last row contains the global test for all the interactions tested at once. Figure 2 shows the plots of log–log survival probabilities against time.

Table 1. Schoenfeld Residuals results for testing PH assumption

Covariate	Rho	Chisq	p-value
Age	-0.0144	0.0456	0.8310
BMI	-0.1114	4.9434	0.0262 **
Sex	0.0617	0.9021	0.0342 **
Family History	-0.0448	0.4530	0.0250 **
Exercise	-0.1461	4.6601	0.0309 **
Alcohol Intake	-0.0969	2.0052	0.0156 **
Salt Intake	0.0195	0.0841	0.7718
GLOBAL	NA	17.1942	0.0162

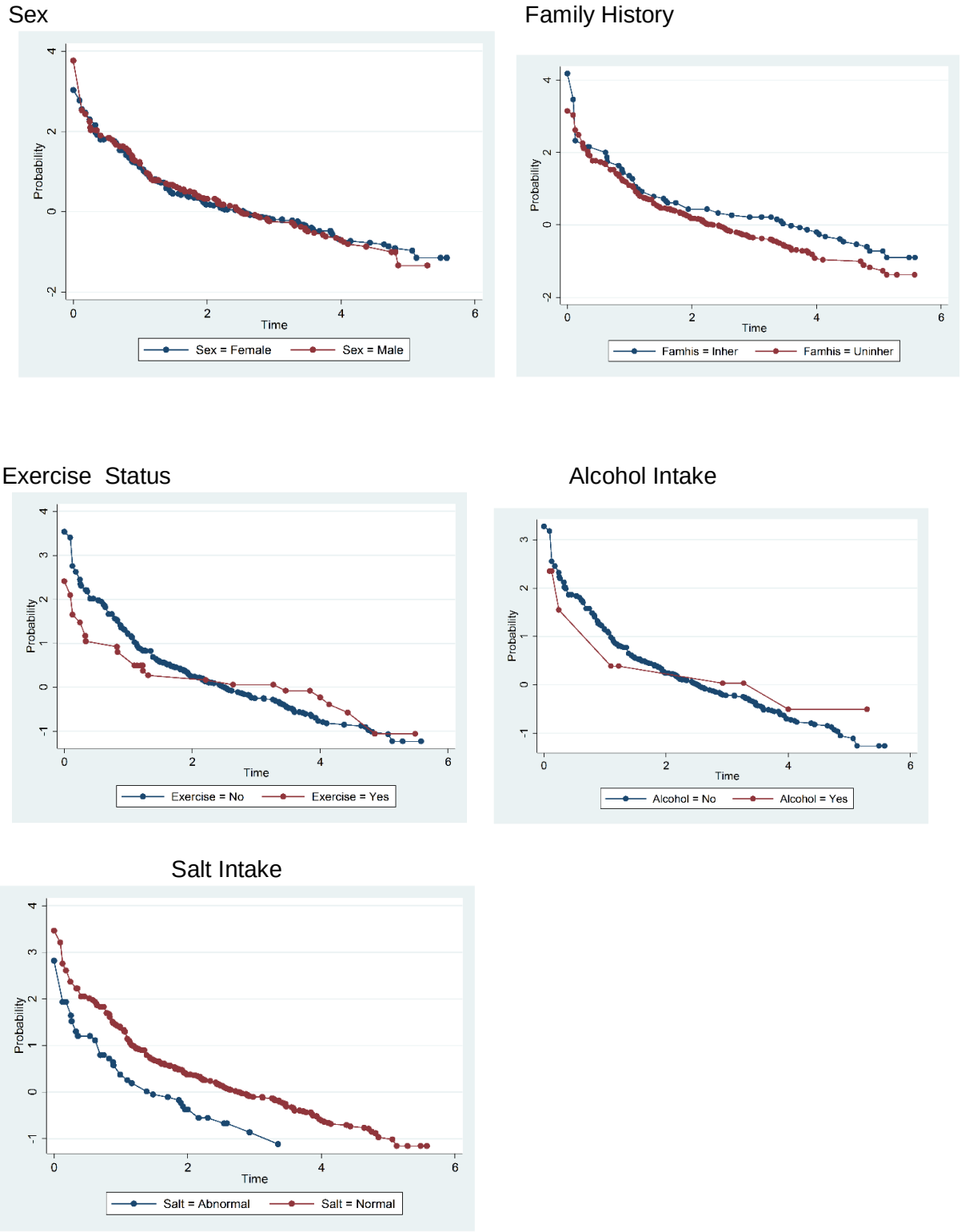


Figure 2. Graphical approach for testing PH assumption

As observed from Table (1) and the graph in Figure 2, PH assumption was violated for categorical covariates sex, family history, exercise status and alcohol as well as continuous covariate body mass index. However, there was no evidence of violation for salt intake (categorical) and age (continuous).The result of global test from the table indicated violation of proportionality assumption (P-value = 0.016).

Results Of Partial And Penalized Partial Likelihoods

The results of Partial and Penalized partial likelihoods are presented in Table (2). As observed, the results are similar for both gamma and inverse Gaussian frailty distributions. However, slightly low standard errors was observed in favour of the penalized partial likelihood estimation method. The significant variables in predicting the risk of death from diabetes were age of the patients, body mass index (BMI), family history and exercise. Age was seen to increase the hazards of death from diabetes across all the configurations, with older patients having higher risk of mortality. For example the estimated hazard ratio (SE), P-value for age are 1.203(0.127), 0.003 under Pl-gam; 1.206(0.123), 0.007 under Ppl-gam; 1.209(0.130), 0.003 under Pl-invgaus and 1.213 (0.127), 0.002 under Ppl-invgaus). It is also observed that body mass index was associated with increased risk of death from diabetes with estimated hazard ratio (SE), P-value for Pl-gam; Ppl-gam; Pl-invgaus; Ppl-invgaus; obtained as 1.073 (0.115), 0.006; 1.043 (0.112), 0.025, 1.068 (0.114), 0.4 and 1.043 (0.113), 0.025 respectively. Family history was positively associated with mortality from diabetes with those having inherited diabetes being more likely to die than those who did not. The estimated hazard ratio (SE), p-value are 1.152 (0.121), 0.002; 1.14 (0.123), 0.004; 1.153 (0.123), 0.002) and 0.154 (0.122), 0.001 for Pl-gam, Ppl-gam, Pl-invgaus and Ppl-invgaus respectively.
Those who engaged in regular exercise had decreased risk of death from diabetes compared to those who did not. As observed from the table, the hazard ratio (SE), p-value = 0.688 (0.067), 0.030 for Pl-gam; 0.597 (0.047), 0.031 for Ppl-gam; 0.686 (0.068), 0.002 for Pl-invgaus and 0.607 (0.047), 0.021 for Ppl-invgaus. It is also observed that these covariates have time-varying effects as evidenced by the p-values of their interactions with function of time (0.0001) being far less than 0.05.

Table 2. Estimates of Hazard Ratio (HR), standard error (SE) and p-value under the four likelihood configurations

Covariate	Pl-gam	PPl-gam	Pl-invgaus	PPl-invgaus
	HR (SE) p-value	HR (SE) p-value	HR (SE) p-value	HR (SE) p-value
Age	1.203 (0.127) 0.003	1.206(0.123) 0.007	1.209 (0.130) 0.003	1.213 (0.127) 0.002
BMI	1.073 (0.115) 0.006	1.043 (0.112) 0.025	1.068 (0.114) 0.004	1.043 (0.113 0.025)
Sex	0.872 (0.078) 0.130	0.814 (0.068) 0.242	0.869 (0.073) 0.156	0.813 (0.064) 0.137
Famhis	1.152 (0.121) 0.002	1.154 (0.121) 0.004	1.153 (0.123) 0.002	1.154 (0.122) 0.001
Exercise	0.688 (0.067) 0.030	0.597 (0.047) 0.031	0.686(0.068) 0.002	0.607 (0.046) 0.021
Alcohol	1.305 (0.136) 0.086	1.306 (0.134) 0.107	1.308 (0.142) 0.261	1.309 (0.140) 0.244
Salt Intake	0.916 (0.108) 0.271	1.016 (0.106) 0.130	0.909(0.110) 0.210	1.101 (0.104) 0.132
BMI*Time	1.225 (0.119) <0.0001	1.208 (0.120) 0.107	1.218 (0.123) 0.107	1.206 (0.122) 0.107
Sex*Time	0.976 (0.081) <0.0001	1.002 (0.079) 0.0001	0.994 (0.088) 0.0001	1.108 (0.082) 0.0001
Famhis*Time	2.242 (0.216) <0.0001	2.235 (0.202) <0.0001	2.244 (0.220) <0.0001	2.233 (0.192) <0.0001
Exercise*Time	0.806 (0.074) <0.0001	0.743 (0.070) <0.0001	0.802 (0.071) <0.0001	0.804 (0.071) <0.0001
Alcohol*Time	1.578 (0.167) <0.0001	1.545 (0.160) <0.0001	1.671(0.173) <0.0001	1.546 (0.154) <0.0001
$\hat{\theta}$	0.503	0.508	0.442	0.439

The estimated frailty variance $\hat{\theta}$ are similar under the two estimation methods, with higher variances for the Gamma frailty distribution compared to Inverse Gaussian distribution. Inclusion of gamma frailty in the model accounted for more subject-specific heterogeneity in the data than the inclusion of Inverse Gaussian frailty.
As observed, the estimated frailty variance ($\hat{\theta}$) for Pl-gam was 0.503 and 0.508 for Ppl-gam, whereas it was 0.442 and 0.439 for Pl-invgaus and Ppl-invgaus respectively.

Model comparison
To compare the methods of estimation as well as the frailty distributions and evaluate the impact of the time-varying covariates, various models were fitted by ignoring and by incorporating time-varying covariates into the models, using the two frailty distributions and under the two estimation methods. Four models, M_1, M_2, M_3 and M_4 were fitted as shown in Table 3, using only the significant covariates that violated the assumption, which are age, body mass index and family history. Model comparison was done using Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) which are given as

AIC = -2loglik+2p
BIC = -2loglik+plog(n),

where loglik is the model loglikelihood, p is the effective number of parameter and n is the number of observations. The model with smaller AIC or BIC is preferred.

Table 3. AIC and BIC for model comparison

Model	Description	AIC		BIC	
		Gam Frailty	**Invgaus Frailty**	**Gam Frailty**	**Invgaus Frailty**
M1	Model with Pl, ignoring time-varying covariates	2314.3	2605.2	2457.8	2676.6
M2	Model with Ppl ,ignoring time-varying covariates	2203.8	2418.6	2344.2	2484.9
M3	Model with Pl incorporating time varying covariates	2018.4	2331.2	2153.7	2395.1
M4	Model with Ppl incorporating time varying covariates	1853.1	2090.5	1983.9	2147.8

As observed from Table 3, comparing the two estimation methods using AIC and BIC, models with Ppl estimation method generally outperformed models with Pl estimation method when time-varying covariates were ignored (M2 versus M1) and when they were incorporated (M4 versus M3). Ignoring time varying covariates adversely affected model adequacy. As observed, models M3 and M4 that incorporated time-varying covariates provided better fits under AIC and BIC criteria than M1 and M2 which ignored them and penalized partial likelihood model with time-varying covariates (M4) was the overall best with AIC of 1853.1 and 2090.5 under Gamma and Inverse Gaussian frailties respectively, while the corresponding BIC values are 1983.9 and 2147.8 respectively. The preference of the frailty distribution was in favour of gamma distribution under both AIC and BIC.

Conclusion
A study on diabetic patients was carried out using Extended Cox model with frailty under Partial and Penalized partial likelihood estimation methods. Proportionality assumptions which underlies the use of Cox proportional hazards model was violated on some covariates, including body mass index (BMI), sex, family history, exercise status and alcohol intake. Thus, an extended Cox model, which is a more appropriate model for time-varying covariates was used. Two frailty distributions, the one-parameter gamma and inverse Gaussian distributions were used to account for unobserved heterogeneity among the diabetic patients. A dichotomous one-step change point Heaviside function was used for the time-varying

covariates. Four likelihood configurations were formed from the combinations of the two estimation methods and frailty distributions. These are Partial likelihood with Gamma frailty, Partial likelihood with Inverse Gaussian frailty, Penalized partial likelihood with Gamma frailty and Penalized partial likelihood with Gamma frailty. It was found that age and body mass index of the patients had significant increased risk of death while regular exercise had significant decreased risk of death from diabetes. Penalized partial likelihood estimation method generally outperformed models with Partial likelihood under all scenarios for the data and Gamma frailty provided a better fit in accounting for unobserved heterogeneity among the diabetic patients.

References

Adeleke, K.A., Abiodun, A. A. & Ipinyomi, R. A. (2015). Semi-Parametric Non-Proportional Hazard Model with Time Varying Covariate. *Journal of Modern Applied Statistical, Methods*, 14 (2), 68-87.

Aro, S., Kangas, T., Reunanen, A., Salinto, M. & Koivist, V. (1994). Hospital use among diabetic patients and the general population. *Diabetes Care*, 17(11):1320–9.

Austin, P.C., Mamdani, M.M., van Walraven C. & Tu, JV. (2006). Quantifying the impact of survivor treatment bias in observational studies. *Journal of Evaluation in Clinical Practice, 12(6), 601–612.*

Austin P.C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. Statistics in Medicine, 31, 3946–3958.

Australian Institute of Health and Welfare. Diabetes prevalence in Australia. An assessment of national data sources. Canberra: AIHW; 2009.

Bellera, C.A,. MacGrogan,G., Debled, M., Christine Tunon de Lara, C. T., Brouste, V. & Mathoulin-Pélissier, S. (2010). Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Medical Research Methodology, 10:20. http://www.biomedcentral.com/1471-2288/10/20.

Beyersmann, J., Wolkewitz, M. & Schumacher M. (2008). The impact of time-dependent bias in Proportional hazards modelling. Statistics in Medicine, 27, 6439–6454.

Bo, S, Ciccone, G., Grassi, G., Gancia, R., Rosato, R., Merletti, F. et al. (2004). Patients with type 2 diabetes had higher rates of hospitalization than the general population. *J Clin Epidemiol*. 57(11):1196–201.

Carral, F., Olveira G, Salas, J., Garcia, L., Sillero, A., Aguilar, M., (2002). Care resource utilization And direct costs incurred by people with diabetes in a Spanish hospital. Diabetes Res & Clin Prac. 56(1):27–34.

Colagiuri, S., Dickinson, S., Girgis, S. & Colagiuri, R. (2009). National Evidence Based Guideline for Blood Glucose Control in Type 2 Diabetes. Canberra. Diabetes Australia and the NHMRC.

Comino, E.J., Harris, M. F., Islam, M. D. F.,Tran, D.T.,Jalaludin, B., Jorm, L., Flack, and Haas, M. (2015). Impact of diabetes on hospital admission and length of stay among a general population aged 45 year or more: a record linkage study. Health Services Research.15:12 DOI 10.1186/s12913-014-0666-2

Cortese, G., Scheike, T., Martinussen, T., (2009). Flexible survival regression modelling. Stat Methods Med Res. 00:1-24.

Cox, D.R. (1972). Regression Models and Life-tables (with discussion). J. Roy. Statist. Soc. Ser. B. 34, 187-220.

De Berardis, G., D'Ettorre, A., Graziano, G., Lucisano, G., Pellegrini, F., Cammarota, S., et al. (2012). The burden of hospitalization related to diabetes mellitus: a population-based study. Nutr Metab & Cardiovasc Dis. 22(7):605–12.

Grambsch, P. & Therneau, T. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika, 81:515-26.

Khalid, J.M, Raluy-Callado M, Curtis BH, Boye KS, Maguire A, Reaney M. (2013). Rates and risk of hospitalisation among patients with type 2 diabetes: retrospective cohort study using the UK General Practice Research Database linked to English Hospital Episode Statistics. Int J Clin Pract., 68(1):40–80.

Kleinbaum, D.G. and Klein, M. (2012). Survival analysis: A Self- Learning Text, Third Edition. Springer New York Dordrecht Heidelberg London.

McGilchrist, C. A. & Aisbett, C.W. (1993). Regression with frailty in survival analysis. Biometrics, 47, 461-466. Munda, M., Rotolo, F. & Legrand. C. (2012). parfm: Parametric Frailty Model in R. Journal of Statistical Software, 15(11), 1-20.

Moolgavkar, S.H., Chang, E. T., Watson, H. N. and Lau, E.C. (2018) . An Assessment of the Cox Proportional Hazards Regression Model for Epidemiologic Studies. Risk Analysis, Vol. 38, No. 4, DOI: 10.1111/risa.12865. 777-794.

Munda, M., Rotolo, F. & Legrand. C. (2012). parfm: Parametric Frailty Model in R. *Journal of Statistical Software*, 15(11), 1-20.

Ng'andu, N.H (1997). An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model. Stat Med. 16:611-26.

Putter, H., Sasako, M., Hartgrink , H.H., van d V, van Houwelingen JC. (2005). Long-term survival With non-proportional hazards: results from the Dutch Gastric Cancer Trial. Stat Med. 24:2807-21.

Rosenthal, M.J., Fajardo, M., Gilmore, S., Morley, J.E. & Naliboff, B.D. (1998). Hospitalization and mortality of diabetes in older adults. A 3-year prospective study. Diabetes Care. 1998;21(2):231–5.

Sertkaya, D., Ata, N., & özer, M. T. (2005). Yaşam cozumlemesinde zamana bağlı acıklayıcı değişkenli Cox regresyon modeli. Ankara Üniversitesi ıp akültesi Mecmuası, 58, 153-58.

Therneau, T. M., Grambsch, P. M., & Pankratz, V. S.(2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12, 156–175.

Suissa, S. (2007). Immortal time bias in pharmacoepidemiology. *American Journal of Epidemiology*, 167(4), 492–499.

Therneau, T. M., Grambsch, P. M., & Pankratz, V. S.(2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12, 156–175.

Vaupel, J.W., Manton, K.G. & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography, 16, 349-454.

Zhou M. (2001). Understanding the Cox regression models with time-change covariates. The American Statistician; 55(2):153{155.DOI:10.1198/000313001750358491.

How to cite this paper:
Alfred, A. A. , Benjamin, A. O. & Kazeem, A. A. (2019). Analysis of Diabetes Data using Extended Cox Model with Frailty under Partial and Penalized Partial Likelihood Estimation Methods. *Malaysian Journal of Applied Sciences, 4*(1), 73-83.