Journal of Modern Applied Statistical Methods

Volume 14 | Issue 2 Article 9

11-1-2015

Semi-Parametric Non-Proportional Hazard Model With Time Varying Covariate

Kazeem A. Adeleke
Obafemi Awolowo University, kay7bright2002@yahoo.com

Alfred A. Abiodun
University of Ilorin, Kwara State, manfredwolw@yahoo.com

R. A. Ipinyomi
University of Ilorin, Kwara State, ipinyomira@yahoo.co.uk

DOI: 10.22237/jmasm/1446350880

Semi-Parametric Non-Proportional Hazard Model With Time Varying Covariate

Cover Page Footnote
I sincerely acknowledge the contributions of Dr Abdul-Raheem AKINDELE, a Lecturer in the department of Psychology, Olabisi Onabanjo University, Ago-Iwoye for his immense contribution and counselling during data collection stage.

Journal of Modern Applied Statistical Methods
November 2015, Vol. 14, No. 2, 68-87.

ISSN 1538 − 9472

Semi-Parametric Non-Proportional Hazard Model with Time Varying Covariate

Kazeem A. Adeleke
Obafemi Awolowo University
Ile-Ife, Nigeria

Alfred A. Abiodun
University of Ilorin, Kwara State
Ilorin, Nigeria

R. A. Ipinyomi
University of Ilorin, Kwara State
Ilorin, Nigeria

The application of survival analysis has extended the importance of statistical methods for time to event data that incorporate time dependent covariates. The Cox proportional hazards model is one such method that is widely used. An extension of the Cox model with time-dependent covariates was adopted when proportionality assumption are violated. The purpose of this study is to validate the model assumption when hazard rate varies with time. This approach is applied to model data on duration of infertility subject to time varying covariate. Validity is assessed by a set of simulation experiments and results indicate that a non proportional hazard model performs well in the phase of violated assumptions of the Cox proportional hazards.

Keywords: Survival time, non-proportional hazards, time-dependent covariate, semi parametric model.

Introduction

In survival or life testing experiments, the assumption of Cox model (1972), may not hold. Example of this is when effect of a treatment on survival diminishes in the course of time to event. Different systems have different prognostic factors, some are time fixed although some are time varying. One advantage of Cox proportional regression models is the ability to incorporate time varying coefficients and time varying covariates (Cox, 1972, Therneau & Grambsch, 2000). The former refers to a variable that is measured at baseline and whose values remain fixed to a variable whose value remains fixed over the duration of follow-up. Although, its effects on hazards is allowed to change over the follow-up period. The later refers to a variable whose value itself varies over time of follow-up. Example of time varying covariate includes the exposure of a pharmaceutical agent to cumulative dosage of radiation, duration of relationship

Kazeem A. Adeleke is a lecturer in the Mathematics Department. Email him at: aadeleke@oauife.edu.ng. Alfred A. Abiodun is a lecturer in the Department of Statistics. R. A. Ipinyomi is an Professor of Statistics. Email him at: ipinyomira@yahoo.co.uk.

as a measure of duration of infertility in marriage, the receipt of an organ transplant. The natures of time varying covariate are very important and take major role of this work. In the above example, the first and second are continuous time variates whose value is non-decreasing over the time, the third example which is the receipt of an organ is also a time varying covariate but dichotomous in nature because the subject may be exposed or unexposed to the treatment.

Recently a number of studies have been directed towards modelling time varying covariates as well as stratification which are semi-parametric non-proportional hazard models (Austin, 2012, Lehr, 2004, Abrahamowicz, 2007, Bender, Augustin, & Blettner, 2005, Ata & Sozer, 2007, Austin, 2012, Zhou, 2001). A more advanced method of generating time varying covariate is the work of Zhou (2001) where the use of an exponential distribution was examined in conjunction with a transformation to the Cox model including time varying covariate. A piecewise exponential distribution was used to obtain a dichotomous or step function covariate which was in turn incorporated into the Cox model and analysed through a semi-parametric approach.

Bender et al. (2005) generated survival data that follows Cox proportional hazard model using three parametric distributions namely: exponential, Weibull and Gompertz and limited his study to only time fixed covariate. New extensions of Cox model with time varying covariate have been developed by Sylvestre and Abrahmowicz (2007) due to an undiscovered and complicated nature of longitudinal data structure where validation is made through simulation. They described and evaluated two alternatives for generation of survival times conditional on time varying covariate.

Applications of Cox model with time varying covariate are likely to continue to become increasingly important in medical research. The methods put forth by Sylvester and Abrahmowicz are however not presented in a close form. Leemis (1987), Leemis, Shih and Ryertson (1990), and Shih and Leemis, (1993) have offered different frameworks for generation of survival time that follow a Cox model with time varying following accelerated life and proportional hazards models where his procedures adopted one time varying covariate and no time fixed covariates. A recent study on Cox regression model in the presence of non-proportional hazards was carried out by Ata and Sozer (2007), where they worked on alternative different models in the violation of proportional assumption. Our study extend the work of Bender et al. (2005), and Zhou (2001), with an additional argument that allows for a fixed covariate, continuous time varying covariate and a step function covariate using exponential model see Austin (2012).

Non-proportional hazards models

Recall the Cox proportional hazards model with time fixed covariate x

$$h_i(t) = h_i(t, \underline{x}) = h_0(t)\exp\left\{\underline{\beta}'\underline{x}\right\} \tag{1}$$

where $h_0(t)$ is a non-parametric baseline hazard function $\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of regression coefficients, and $x_i = x_1, x_2, \ldots, x_p$ is a vector of time fixed covariates for ith subject.

Although $h_0(t)$ is chosen arbitrarily with no distribution attached, the fact that $\exp\left(\underline{\beta}'\underline{x}\right)$ is a parametric exponential function that assumes parametric forms of the predictors on hazards makes model in (1) a semi-parametric model.

Proportional hazard assumption

In linear regression modelling, the measure of effect is usually regression coefficient β, in logistic regression the measure of effect is an odds ratio, Walker and Duncan(1976), Hosmer and Lemeshow (2000), Agresti (2007), Adeleke and Adepoju (2010), the log of which is β, but in survival analysis, the measure of effect is the hazard ratio (Tableman and Kim, 2004). Proportional hazards assumption states that the hazard ratio is constant over time or the hazard for an individual is proportional to the hazard for any other individual (Kleinbaum and Klein, 2005). For example, if x and x^* are the covariates for two individual then

$$HR = \frac{h\left(t\middle|x\right)}{h\left(t\middle|x^*\right)} = \frac{h_0(t)\exp\left(\underline{\hat{\beta}}'\underline{x}\right)}{h_0(t)\exp\left(\underline{\hat{\beta}}'\underline{x}_*\right)} = \exp^{\left(\sum(\underline{x}-\underline{x}_*)'\underline{\hat{\beta}}\right)} \tag{2}$$

The hazard ratio in (2) can also be expressed as $HR = \theta$, which implies that the hazard ration is time-independent.

Now let the effect of a time varying covariate on survival probability at a time $t(\beta_t)$ depend on the value of this variable at the same time, then an extended version of (1) by Cox (1972) can be given by

$$h\left(t, z(t), x\right) = h_0(t)\exp\sum_1^{p1}\beta_i x_i + \sum_1^{p2}\gamma_i z_i(t) \tag{3}$$

which can be written as $HR = \theta t$

Let the proportional hazard for a survival time T be given by

$$h(T_i / X) = \exp\left(\underline{\beta}'\underline{x}\right) h_0(t) \tag{4}$$

Then the cumulative distribution of T_i can be given as

$$F_T(t) = P(T_i \le t) = P\left(\exp(\beta' x) h_0(t) \le t\right) \tag{5}$$

$$\begin{aligned} H(t, z(t), x) &= \lambda \exp(\beta' x) t \\ F_T(t) &= P(T_i \le t) \\ &= P\left(\exp\left(\underline{\beta}'\underline{x}\right) h_0(t) \le t\right) \end{aligned} \tag{6}$$

$$\begin{aligned} P\left(h_0(t)\right) &\le \frac{t}{\exp\left(\underline{\beta}'\underline{x}\right)} \\ &= F_{\exp(1)}\left(\frac{t}{\exp\left(\underline{\beta}'\underline{x}\right)}\right) \\ &= 1 - \exp\left(-\frac{t}{\exp\left(\underline{\beta}'\underline{x}\right)}\right) \end{aligned}$$

Now if $S_T(t) = 1 - F_T(t)$

$$T_i \approx S_T(t) = \exp\left(-\frac{1}{\exp\left(\underline{\beta}'\underline{x}\right)}\right) \tag{7}$$

Let Y_i be a uniform random variable with cumulative distribution function F and density function f, then

$$U = F(Y_i) \sim U(0,1)$$
$$F(Y_i) = \int_{-\infty}^{y} f(u) \sim U(0,1)$$

Also

$$U = \exp\left[-H(T)\exp\left\{\underline{\beta}'\underline{x}\right\}\right] \sim U(0,1)$$

$$T = H^{-1}\left[-\log(U)\exp\left(\underline{\beta}'\underline{x}\right)\right] \tag{8}$$

where U is a uniform random variable (Bender et al, 2005). However, the survival time T does not involve time varying variable(s). By introducing the second covariate with time change when covariate is dichotomous, following the formulation of Zhou (2001) and Austin (2012), we define

$$Z_i(t) = \begin{cases} 0, & \text{for } t \le t_0 \\ 1, & \text{fot } t > t_0 \end{cases}$$

then the hazard function with dichotomous time changed covariate is

$$h_g(Y_i) = h_0(t)\exp\left\{\underline{\beta}'\underline{x} + \gamma'\underline{z}(t)\right\} \tag{9}$$

A natural problem is when time varying covariate is not dichotomous or step function but continuous. Zhou (2001) did not consider this, and Sylvestre and Abrahamowicz (2007) found the method was limited in applicability. For a case open to both time fixed and time varying covariate which is flexible for both step function and continuous system, see Austin (2012).

The cumulative hazard function and survival function H(.) and S(.) are:

$$H\left(t\,|\,z(t),x\right) = \int_0^t h_0(s)\exp\left(\underline{\beta}'\underline{x} + \gamma'\underline{z}(s)\right)ds \tag{10}$$

$$S\left(t,z(t),x\right) = \exp\left[-H\left(t,z(t),x\right)\right]$$

Suppose the covariate follows a step function for $t \ge t_0$ i.e right censored data, then supposed the time is partitioned into two such that

$$Z_i(t)=\begin{cases}0, & \text{for } t<t_0\\ 1, & \text{fot } t\geq t_0\end{cases}$$

Let D = domain and $D_1 = [0, t_0)$ and $D_2 = [t_0, \infty)$ then,
for $t < t_0$,

$$\begin{aligned}H\left(t,z(t),x\right)&=\int_0^t \lambda\exp\left(\underline{\beta}'\underline{x}+\gamma'\underline{z}(s)\right)ds\\ &=\lambda\exp\left(\underline{\beta}'\underline{x}\right)\int_0^t ds\\ &=\lambda\exp\left(\underline{\beta}'\underline{x}\right)[s]_0^t\\ &=\lambda\exp\left(\underline{\beta}'\underline{x}\right)t\\ H\left(t,z(t),x\right)&=\lambda\exp\left(\underline{\beta}'\underline{x}\right)t\end{aligned} \tag{11}$$

Using Bender et al. (2005), we obtain survival time

$$T=\frac{-\log U}{\lambda\exp\left(\underline{\beta}'\underline{x}\right)} \tag{12}$$

By Austin (2012), when $t \geq t_0$, using the condition above, the hazard function in (9) becomes

When $D_2 = t \geq t_0$, from 5, $Z(u) = 1$ then 6 becomes

$$\begin{aligned}&=\int_0^{t_0}\lambda\exp\left(\underline{\beta}'\underline{x}\right)du+\int_{t_0}^{0}\lambda\exp\left(\underline{\beta}'\underline{x}+\gamma\right)du\\ &=\lambda\exp\left(\underline{\beta}'\underline{x}\right)t_0+\lambda\exp\left(\underline{\beta}'\underline{x}+\gamma\right)\left(t-t_0\right)\\ &=\lambda\exp\left(\underline{\beta}'\underline{x}\right)t_0\left[1-\exp(\gamma)\right]+\lambda\exp\left(\underline{\beta}'\underline{x}+\gamma\right)t\end{aligned}$$

by transformation

$$-\log(U)=\lambda\exp\left(\underline{\beta}'\underline{x}\right)t_0\left[1-\exp(\gamma)\right]+\lambda\exp\left(\underline{\beta}'\underline{x}+\gamma\right)T$$

The survival time obtained from the inverse cumulative hazards is

$$T = \left[\frac{-\log(U) - \lambda \exp\left(\underline{\beta}'\underline{x}\right)\left[t_0 + \lambda \exp(\gamma_i)(t - t_0)\right)}{\lambda \exp\left(\beta_i x_i + \gamma_i\right)} \right] \tag{13}$$

If however covariate is continuous the cumulative hazards is

$$H\left(t \middle| z(t), x\right) = \int_0^t h_0(s) \exp\left\{\underline{\beta}'\underline{x} + \gamma'\underline{z}(s)\right\} ds \tag{14}$$

Assume that $\underline{z}(s)$ is proportional to t such that $\underline{z}(s) = kt$ where $k > 0$. Hence the cumulative hazard from the above becomes

$$\begin{aligned} H\left(t \middle| z(t), x\right) &= \int_0^t \lambda \exp\left\{\underline{\beta}'\underline{x} + \gamma'\underline{z}(s)\right\} ds \\ &= \lambda \exp\left(\underline{\beta}'\underline{x}\right) \int_0^t \exp(\gamma k s) ds \\ (t, k, x) &= \frac{\lambda \exp\left(\underline{\beta}'\underline{x}\right)}{\lambda k} \left[\exp(\gamma k t - 1)\right] \end{aligned} \tag{15}$$

Hence

$$U = \exp\left(-H(T, k, x)\right)$$

$$\left(-\log(U)\right)\gamma_i k = \lambda \exp\left(\underline{\beta}'\underline{x}\right)\left[\exp(\gamma_i k T - 1)\right]$$

so that

$$T = \frac{1}{\gamma_i k} \log \frac{\left[\lambda \exp\left(\underline{\beta}'\underline{x}\right) + \gamma_i k\left(-\log(U)\right)\right]}{\lambda \exp\left(\underline{\beta}'\underline{x}\right)} \tag{16}$$

Equations (12) and (13) and (16) will be used to obtain survival times for dicotonomous time varying covariate and continuous time varying; *U* can be obtained from *R*.

Non-parametric estimation

Follow the formulation of Kaplan and Meier (K-M) (1958) for estimating censored data. The method provides alternative way to life table approach where each interval contains only one observation.

The idea of K-M estimator is given by the conditional probability ($t \leq t_0$) be the survival time of n randomly sampled individual study such that $t_1 \leq t_2 \leq ,.., \leq t_n$ are of T_1, T_2, ..., T_n where $S(t) \sim b(n, p)$ and $P = P(T \geq t)$ then, for $t \leq t_{i+1}$

$$\begin{aligned} S(t_i) &\leq P(T > t_i, T > t_{i-1}) \\ &= P(T > t_i | T > t_{i-1}) P(T > t_{i-1}) \\ &= P(T > t_i | T > t_{i-1}) P(T > t_{i-1} | T > t_{i-1}) \ldots P(T > t_0 = 0) \end{aligned}$$

Assume that at the beginning of the study all subjects were alive so, $P(T > t_0 = 0) = 1$, and

$$P(T > t_i | T > t_{i-1}) = \frac{n_i - d_i}{n_i}$$

The Kaplan Meier estimator is

$$S_{KM}(t) = \prod i : t_i \left(\frac{n_i - d_i}{n_i} \right) \text{ or } \prod i : t_i \leq \left(1 - \frac{d_i}{n_i} \right)$$

For detail, see Greenwood (1926), Kaplan and Meier (1958), Adeleke (2012).

Semi-parametric estimation

For proportional hazard model of equation (1) where $h_0(t)$ is non-distributional and $\exp(\underline{\beta}' \underline{x})$ is a parametric function, we use partial likelihood estimate of Cox (1975)

$$L_i\left(\underline{\beta}\right) = \prod_{j=1} L_i\left(\underline{\beta}\right)$$
$$= \prod_{j=1} \frac{\exp\left(\underline{\beta}'\underline{x}_j + \gamma'\underline{z}_j\left(t\right)\right)}{\sum_{i\in R} \mathrm{e}\,xp\left(\underline{\beta}'\underline{x}_i + \gamma'\underline{z}_i\left(t\right)\right)}$$

Application to a data of infertility

Data on period of infertility among women were obtained from a survey conducted in 2011 at Ijebu North Local Government (INLG) area of Ogun state. Information on the duration of infertility in years before a woman to get pregnant together with the causes of infertility were collected, along with covariates: duration of relationship (drelation) in years, respondent's age in years, marital status (married, cohabiting and single) and previous infertility treatment such as (ovulation induction, tubal surgery, antibiotic for infection, intercourse during fertile period and assisted conception).

Duration of infertility was measured as the time from marriage/first date of diagnose till fertile/date of first conception or the end of the study.

Let $\delta_i = 1$ if a woman $i = 1, 2, \ldots, n$ become fertile at time t_i and $\delta_i = 0$, if otherwise; let the survival time $T = \min(t_i, C_i)$, where t_i is the observed time and C_i is the censored time. Censored if either lost to follow-up or does not observe the event of interest (get pregnant) within the period of follow-up. First, consider the model of eqn (1) where age and duration of relationship and others were considered to be time fixed. The estimated regression coefficients are given in Table 1 together with associated *p*-values and Schoenfeld test result. As observed, intensity of being fertile is much higher for previous infertility treatment using ovulation induction and antibiotic for treatment of infections than when assisted with conception. Almost all the factors are negatively related with the hazards for the period of infertility. The aim is to know if model (1) is better used for the data or model 3 (i.e whether PH model assumption is satisfied or not). Age and duration of relationship were found to be significant.

Table 2 gives the estimates when age and duration of relationship are categorized as 1 if age less than 19 years, i.e (1-18), 2 if between (19-35) years inclusive and 3 if greater than 35 years. The result is not different much from what we had in Table 1. An indication of a significant variable implies the possibility of the variable varying with time and that implies violation of PH model assumption subject to some tests. The last column of the table is a report

from Schoenfeld test with their respective *p*-values. The *p*-values for the correlation coefficient between time and covariates (duration of relationship) shows a significant relationship, supported by the Schoenfeld plot see fig 2. Another graphical test is log cumulative hazard plot. Log-cumulative hazard curves in fig 1 shows that only age of mothers is violating the assumption. Following the numerical test of the correlation coefficient between variable age of mothers and duration of relationship and time in Table 3, the *p*-values for both coefficients and Schoenfeld residual test for age of mothers and duration of relationship with time are indication that both age of mothers and duration of relationship are time varying.

Having detected this, an extended version of model (1) (i.e model 3) was introduced with age and duration of relationship categorized to see the effect within the age group (0-18, 19-34 and above 35) as shown in Table 4. Here the model is stable with the global test of Schoenfeld test showing a sign of proportionality.

Next, compare the two models, using Akaike's information criterion (AIC) or -2loglikelihood function (-2loglik). The values of AIC and -2loglik for Cox regression and Extended Cox are given in Table 5. According to the results, Extended Cox model gives most suitable result for modelling time to infertility data in the presence of non-proportional hazards followed by Cox model.

Results from infertility data

Table 1. Result from Cox model with Age, duration of relationship continuous

Variables	β (*p*-value)	Schoenfeld Test (rho))(*p*-value
Age	-0.086(1.4e-05)	0.169(0.198)
married	-1.67(0.108)	-0.024(0.840)
Cohabiting	-18.0(0.996)	-0.004(1.000)
drelation	-0.065(0.007)	0.287(0.028)
Ovulation	0.680(0.503)	0.066(0.591)
Tubla.S	-18.2 (0.998)	0.112(0.999)
Antibiotic	0.401 (0.697)	0.021(0.859)
Intercourse	-0.626 (0.659)	0.110(0.356)
		Global (0.0368)

Table 2. Result from Cox model with Age, duration of relationship categorized.

Variables	β (p-value)	Schoenfeld Test (rho)(p-value)
Age<=18	-0.777(0.460)	0.083(0.52)
Age>35	-1.225(4.30E-05)	-0.006(0.956)
Married	-1.69(0.103)	-0.011(0.93)
cohabiting	-17.827(1.00)	0.031(1.00)
dlv.cat1	0.447(0.110)	-0.201(0.0146)
Ovulation	0.862(0.400)	0.068(0.584)
Tubla.S	-17.448(1.00)	0.127(1.00)
Antibiotic	0.49(0.630)	0.026(0.584)
intercourse	-0.38(0.790)	0.087(0.479)
		Global (0.0506)

Table 3. Test for age and duration of relationship as time varying covariates

Variables	β (p-value)	Schoenfeld Test (rho)(p-value)
married	-1.0271(0.320)	-0.053(0.661)
cohabiting	-17.277(1.000)	-0.053(1.00)
Ovulation	0.94(0.360)	0.031(0. 802)
Tubla.S	-18.594(1.000)	0.086(1.00)
Antibiotic	0.617(0.550)	0.003(0. 980)
intercourse	-0.638(0.650)	0.130(0.283)
Age* time	-0.0187(0.000)	0.613(1.23E-09)
Drelation*time	-0.0055(0.021)	0.295(3.16E-03)
		Global(1.41E-06)

Table 4. Extended Cox model with age and duration of relationship as time varying.

Variables	β (p-value)	Schoenfeld Test (rho)(p-value)
married	-0.986(0.340)	0.0128(0.918)
cohabiting	-6.713(0.760)	0.0006(1.00)
Ovulation	1.384(0.180)	0.078(0.528)
Tubla.S	-8.640(0.940)	0.136(0.988)
Antibiotic	1.0257(0.320)	0.049(0.689)
intercourse	-0.275(0.840)	0.12908(0.296)
Age<=18*time	-4.612(1.70E-06)	0.194(0.548)
age.cat2*time	-4.717(1.0E06)	0.183(0.56)
Age>35*time	-4.713(9.70E-07)	0.198(0.544)
Time*dlv.cat1	0.001(0.980)	0.102(0.366)
		Global (0.982)

Table 5. AIC and -2loglik values.

	PHM	NPHM Extended Cox
AIC	525.813	311.6885
Loglik	509.813	291.688

Results from Simulation

Table 6. Mean values of the estimated regression coefficients for continuous time varying covariate model (16).

% cens	$\hat{\beta}$	$\hat{\gamma}$	AIC	loglik
C=0.0	-0.849(0.007)	0.724(0.151)	473.392	-309.929
C=0.5	-0.976(0.112)	2.016(0.0003)	158.962	-105.449
C=0.8	-0.770(0.261)	2.389(0.049)	62.032	-50.788

Table 7. Sample variances of the estimated regression coefficients for continuous time varying covariate model (16).

% cens	$\hat{\beta}$	$\hat{\gamma}$
C=0.0	0.0619	0.0552
C=0.5	0.1793	0.2073
C=0.8	0.4580	0.5744

Table 8. Mean values of the estimated regression coefficients for dicotonomous time varying ($t \geq t_0$); model 13.

% cens	$\hat{\beta}$	$\hat{\gamma}$	AIC	loglik
C=0.0	-0.363(0.211)	0.299(0.238)	625.857	-233.696
C=0.5	-0.348(0.363)	0.692(0.201)	240.578	-75.969
C=0.8	-0.184(0.411)	0.572(0.313)	107.576	-28.016

Table 9. Sample variances of the estimated regression coefficients for dicotonomous time varying ($t \geq t_0$); model 13.

% cens	$\hat{\beta}$	$\hat{\gamma}$
C=0.0	0.0537	0.0457
C=0.5	0.1271	0.1132
C=0.8	0.2664	0.2086

Table 10. Mean values of the estimated regression coefficients for time fixed covariate ($t \geq t_0$); model 12.

% cens	$\hat{\beta}$	$\hat{\gamma}$	AIC	loglik
C=0.0	-0.998 (2e-16)	0.043 (0.165)	11619.89	-5807.947
C=0.5	-1.058 (2e-16)	2.152 (2e-16)	5313.93	-2654.965
C=0.8	-8.060(2.4e-15)	-1.94(2e-16)	2585.184	-1290.592

Table 11. Sample variances of the estimated regression coefficients time fixed covariate ($t \geq t_0$); model 12.

% cens	$\hat{\beta}$	$\hat{\gamma}$
C=0.0	0.0047	0.00097
C=0.5	0.0088	0.0061
C=0.8	1.0365	0.0114

Table 12. Absolute Bias continuous TVC model 16**.**

% cens	$\underline{\beta}$	Abs Bias	MSE
$C = 0.0$	$\beta = -1$	0.150	0.069
	$\gamma = 0$	0.723	0.751
$C = 0.5$	$\beta = -1$	0.023	0.201
	$\gamma = 2$	0.015	0.257
$C = 0.8$	$\beta = -1$	0.229	0.659
	$\gamma = 3$	0.611	1.465

Table 13. Absolute Bias for dicotonomous time varying ($t \geq t_0$); model 13.

% cens	$\underline{\beta}$	Abs Bias	MSE
$C = 0.0$	$\beta = -1$	0.636	0.471
	$\gamma = 0$	0.298	0.143
$C = 0.5$	$\beta = -1$	0.651	0.611
	$\gamma = 2$	1.308	1.994
$C = 0.8$	$\beta = -1$	0.815	0.996
	$\gamma = 3$	2.428	6.143

Table 14. Absolute Bias for time fixed covariate ($t \geq t_0$); model 12.

% cens	$\underline{\beta}$	Abs Bias	MSE
$C = 0.0$	$\beta = -1$	0.002	0.918
	$\gamma = 0$	0.043	0.211
$C = 0.5$	$\beta = -1$	0.058	0.221
	$\gamma = 2$	0.152	1.133
$C = 0.8$	$\beta = -1$	7.06	1.110
	$\gamma = 3$	4.94	2.720

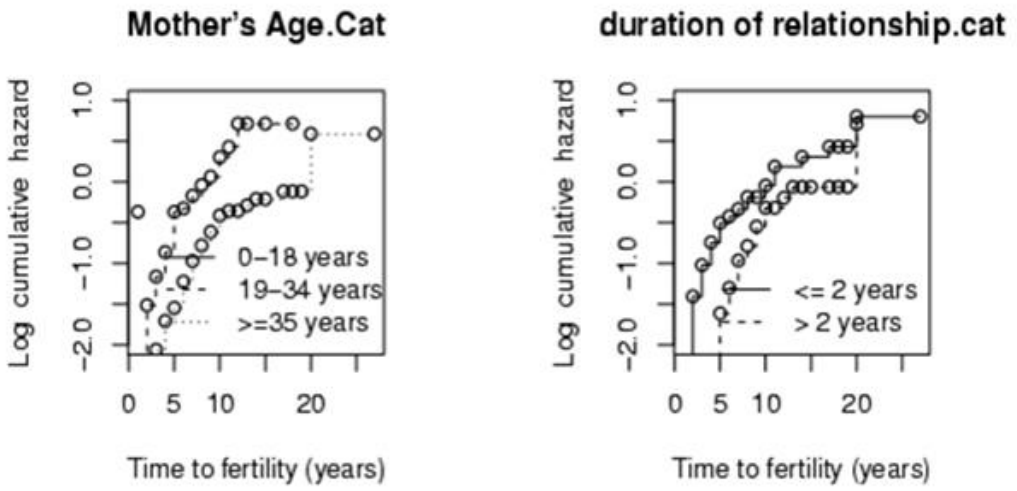

Figure 1. Log cumulative hazards for age and duration of relationship.

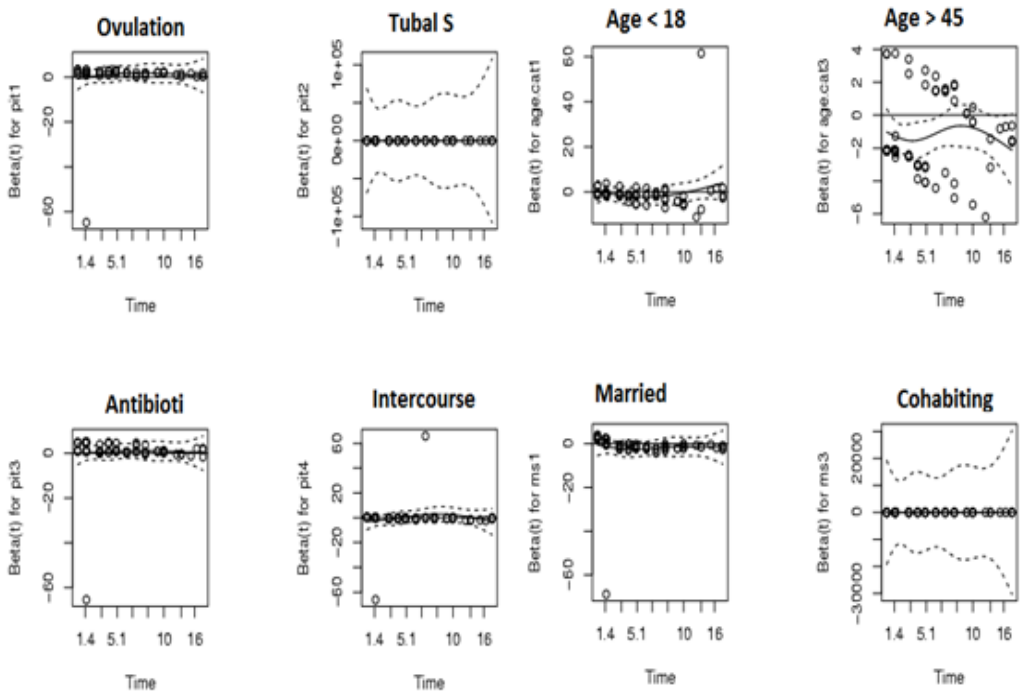

Figure 2. Schoenfeld Plots of residuals

In purpose of the simulation was to investigate the violation of the assumption and the use of Non-proportional hazard Model for different values of the true parameters β and γ, at different level of censoring. Hypothesis about the regression coefficients β and γ of the model 1.0 in various situations was tested. Each simulation consists of 80 replicates. The set-up of the simulated data resembles that of right censored and truncated data. For each sample, 1000 samples of survival times (months) were generated.

Given a time t_* , the time u were generated from a uniform $(0, t_*)$ distribution although the baseline survival time t_i were generated from an exponential distribution for fixed and time varying covariates in term of continuous and dichotomous covariates as define in eqn 12, 13 and 16. Two covariates; a time fixed and a binary with $P(z = 0) = P(z = 1) = ½$ and the other is distributed as normal and varies with time. Only the data that satisfy the condition $u_i + t_i \leq t_*$ were kept in the sample given rise to right truncated data. The survival time is not only right truncated but also right censored. The simulation was carried out at three different percentage of censoring viz: 0%, 50% and 80%.

The true values of regression coefficients β, γ were taken to be either (-1, 0), (-1, 2), (-1, 3) in the simulation each at different level or percentage of censoring. Comparison were made using absolute bias Tables 6 to 11 showed the estimated mean values of $\hat{\beta}$ and $\hat{\gamma}$, p-values as well as the sample variances. The result in Tables 6 and 9 are from the analysis of (3) through the use of survival time obtained in (16) for fixed and continuous time varying covariates of (3). The estimated coefficients $\hat{\beta}$ is for the fixed covariate although $\hat{\gamma}$ is for the time varying (continuous or binary). The coefficients are significant at 50% and 80 % censoring and slightly overestimate its true value as percentage of censoring increases resulting in higher variance than the estimator of the other coefficient which appear to be more stable with lower variance than γ. Absolute Bias (AB) of Tables 12 to 14 showed the sensitivity of the model to change in percentage of censoring. At 0 percent censoring, model with time fixed covariate has the minimum AB followed by model with continuous time varying covariate. Also at 50% censoring, model with continuous time varying covariate has the minimum AB, followed by model with time fixed covariate. At 80% censoring, model with continuous time varying covariate has the minimum AB next is model with dicotonomous time varying covariate and least is time fixed model.

Checking the parameter of the time varying coefficient, as the values of the parameter γ increases from 0 to 3, At $\gamma = 0$, the AB of the parameter is minimum for model with time fixed covariate, followed by a model with dicotonomous time varying covariate and maximum for model with continuous time varying covariate. At $\gamma = 2$, AB is minimum for semi-parametric model via continuous time varying covariate (model 16), followed by a time fixed and maximum for semi-parametric model with dicotonomous time varying covariate. Lastly at $\gamma = 3$ AB increases from model with continuous time varying covariate to semi-parametric model with time fixed covariate. Hence, as parameter of time varying coefficient increase from 0-3, the semi-parametric model with continuous time varying covariate showed the minimum AB followed by dicotonomous time varying covariate and maximum with time fixed covariate model. This actually showed an evidence of time varying both in the coefficient and covariate.

For Mean Square Error (MSE), Semi-parametric model with continuous time varying covariate has being the best (with min MSE) among the three models as percentage of censoring increases from 0% to 80 percent. Also as parameter of time varying coefficient increases from 0 to 3, parameters of the semi-parametric model with continuous time varying coefficient showed the minimum MSE, and perform best. Followed by the parameters of time fixed

covariate model and maximum MSE with model with dicotonomous time varying covariate.

Discussion

The result is more encouraging at 80% of censoring resulting from the outcome of the AIC and log-likelihood estimates of model selection criteria and generally accepted for all other results. Percentage of censoring contributes to the outcome and conclusion in that as the level of censoring increases from 0% through 50% to 80%. The coefficients of time varying covariates varying from zero to three (0-3). See Tables 6 and 10, the result also give a good sign of a well satisfactory size and power. The higher the percentage of censoring, the more closely the violation of PHM. It implies that at 80% censoring which is generally accepted from the results of our simulated data there exist an outright violation of the assumption of proportionality and this assume a semi-parametric non proportional hazard model.

In Tables 8, 9, 10, and 11 models 12 and 13 were used to generate survival time when both covariates are dichotomous and continuous, although time varying. The time varying covariate $Z(t)$ is zero when $t < t_0$ and 1 when $t \geq t_0$ as stated in the model, our t_0 is the maximum time it takes a woman to conceive (i.e 24 months), see Esther, Eunice , Kelly, CHESRenee, and Lee (2009), Ekwere, et al (2007) and Yusuff (2006). (When $t < t_0$, we obtain our survival time as we have in (12) and when $t \geq t_0$, it resulted in survival time of (13) as we notice from the estimated mean values and variances of Tables 8 and 9. None of the coefficients at any level of censoring is significant judging from the PH values of the coefficient. An indication of satisfying PH model assumption, but when $t \geq t_0$ (dicotonomous), the estimated mean values and sample variances of regression coefficient does not satisfy PH model assumption following parameters significant properties of the coefficients from the *p*-values.

The model with continuous time varying covariate (model 16) performed better (min AB and MSE) followed by model with dicotonomous time varying covariate and least with model with time fixed covariate see Tables 12 to 14. The same result follows when parameters of the time varying coefficient increase from 0-3.

References

Abrahamowicz, M., & MacKenzie, T. A. (2007). Joint estimation of time-dependent and non-linear effects of continuous covariates on survival. *Statistics in Medicine, 26*(2), 392-408. doi:10.1002/sim.2519

Adeleke, K. A. (2012). A simulation study on kaplan meier non-parametric survival methods. *Journal of the Nigerian Mathematical Society, 31*, 243-254.

Adeleke, K.A. & Adepoju, A. A. (2010). Ordinal logistic regression model: An application to pregnancy outcomes. *Journal of Mathematics and Statistics, 6*(3), 279-285. doi:10.3844/jmssp.2010.279.285

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: Wiley.

Ata, N. & Sozer, M. T. (2007). Cox regression models with non-proportional hazards applied to lung cancer survival data. *Hacettepe Journal of Mathematics and Statistics, 36*(2), 157-167.

Austin, P. C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine, 31*(29), 3946–3958.doi:10.1002/sim.5452

Bender. R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine, 24*(11), 1713-1723. doi:10.1002/sim.2059

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological), 34*(2), 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika, 62*(2), 269-276. doi:10.2307/2335362

Ekwere, P. D., Archibong, E. I., Bassey, E. E., Ekabua, J. E., Ekanem, E. I. & Feyi-Waboso, P. (2007). Infertility among Nigerian couples as seen in Calabar. *Port Harcourt Medical Journal; 2*(1), 35-40. doi:10.4314/phmedj.v2i1.38890

Eisenberg, E., Shriver, E. K., Brumbaugh, K., Brown-Bryant, R & Warner, L. (2009). Frequently asked questions. *United State Department of Health and Human Services, Office on Women's Health*. Retrieved from http://www.womenshealth.gov

Greenwood, M. (1926). A report on the natural duration of cancer. *Reports on Public Health and Medical Subjects. Ministry of Health*, (33).

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457-481. doi:10.2307/2281868

Kleinbaum, D. G. & Klein, M. (2005). *Survival analysis: A self learning text* (2nd ed.). New York: Springer-Verlag.

Leemis, L. M. (1987). Techinical note-Variate generation for accelerated life and proportional hazards models. *Operations Research, 35*(6), 892-894. doi:10.1287/opre.35.6.892

Leemis, L. M., Shih, L. H., & Reynertson, K. (1990). Variate generation for accelerated life and proportional hazards models with time dependent covariates. *Statistics and Probability Letters; 10*(6), 335-339. doi:10.1016/0167-7152(90)90052-9

Lehr, S. (2004). *Over-fit in the analysis of time-dependent effects of prognostic factors.* Ph.D. Thesis, University of Vienna.

Shih, L. H. & Leemis, L. M. (1993). Variate generation for a nonhomogenous poisson process with time dependent covariates. *Journal of Statistical Computation and Simulation*; *44*(3-4), 165-186. doi:10.1080/00949659308811457

Sylvestre, M. P. & Abrahamowicz, M. (2007). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine; 27*(14), 2618-2634.doi:10.1002/sim.3092

Tableman, M. & Kim, J. S. (2004). *Survival analysis using S: Analysis of time-to-event data.* New York: Chapman and Hall/CRC.

Therneau, T. M. & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer-Verlag.

Walker, S. H & Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika, 54*(1-2), 167-179. doi:10.1093/biomet/54.1-2.167

Yusuff, A. O. (2006). Infertility treatments and the conceptual dilemma of infertility: is there a right to reproduce? *Canadian Institutes of Health Research (CIHR) at the University of Toronto, ON, Canada.* Retrieved from http://www.nigerianlawguru.com/articles/human%20rights%20law/INFERTILITY%20TREATMENTS%20AND%20THE%20CONCEPTUAL%20DILEMMA%20OF%20INFERTILITY,%20IS%20THERE%20A%20RIGHT%20TO%20REPRODUCE.pdf.

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