INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

CPH and AFT Models for Time-To-Employment Data in the Presence of Cure Fraction

Zahraddeen Abdullahi, Zarina Mohd Khalid*, Haliza Abd Rahman, Nur Arina Bazilah Kamisan

Department of Mathematical Sciences, Faculty of Science, University Technology Malaysia, Malaysia

***Corresponding author**

DOI : https://dx.doi.org/10.47772/IJRISS.2024.8120004

Received: 15 November 2024; Accepted: 19 November 2024; Published: 26 December 2024

ABSTRACT

In this work, we present the use of mixture cure models (MCM) to analyze time-to-employment data of graduates of the statistics department, Kano University of Science and Technology, Nigeria. This is against Cox proportional hazards (CPH) and accelerated failure time (AFT) models that are traditionally used to model such types of data. MCM was used because the Kaplan-Meier (KM) employment curve has suggested the possibility of cure with an estimated unemployment fraction of 33.8%. Here, two MCM were constructed based on CPH and AFT assumptions for the latency part of the model. Weibull was used as the baseline distribution in the AFT Cure model. The Cure models were used to estimate the unemployment fraction, survival function of the employment subgroup, as well as the effects of covariates on time-to-employment and probability of unemployment. Estimates of unemployment fractions by CPH Cure model are closer to the empirical estimate by KM compared to that of Weibull AFT Cure model. In comparing cure and non-cure CPH, some of the covariates (Gender and Age) that were not significant in the non-cure model were found to be significant in the cure model. Likewise Grade, which was found to be significant in the non-cure model, was not significant in the cure model. None of the covariates was found to influence unemployment probability significantly. It is concluded that, since today’s time-to-employment data of graduates mostly consists of groups that would remain unemployed forever (cure fraction), then the use of cure models is superior to their non-cure counterparts in revealing the true effect and significance of a covariate on time-to-employment. In addition, cure models assess the influence of covariates on unemployment probability. Findings may benefit the government and other stakeholders in employment planning policies.

Keywords: Cure fraction, mixture cure model, CPH, AFT, employment

INTRODUCTION

The study of graduate employability after graduation has for long being shifted from merely use of techniques like descriptive statistics, linear or logistic regressions to a more appropriate technique of survival analysis (Barros, Guironnet and Peypoch, 2010; Lim, 2008; Manuel, 2007; Rosna et al.,2015). Here, time-to-employment is modelled statistically as a response variable that depends on some graduates’ characteristics, as such emphasis is much on identifying those characteristics that influence the transition to employment. Some of the factors that influence time-to-employment includes gender, age at graduation, grade/class of certificate/CGPA, course of study and Marital status among others (Ayaneh, Dessie and Ayele, 2020; Ezeani, 2018; Soon, Lee, Lim, Idris and Eng, 2019).

If modelling survival data involves estimating the effects of covariates, Cox Proportional Hazard (CPH) (Cox, 1972) and Accelerated Failure Time (AFT) survival regression models were mostly used. AFT models are parametric models and are more appropriate when the time-to-event of interest can be appropriately described by a parametric distribution otherwise CPH model is the best choice. CPH model is very popular because it does not assume any distribution for the time-to-event of interest and yet is robust in its estimate, only that its appropriateness largely depends on satisfying the so-called proportional hazard assumption.

The transition period from graduation to first employment is increasing by the day and more concern is the frequent changes in the transition pattern compared to smooth transition in the past (Ayaneh et.al, 2020). As time goes by from graduation before employment, graduates do deflate in their human capital or skills they acquire in universities. This in turn reduces their chances of getting job. Now that unemployment rate is increasing (FGN, 2017; Lim, Rich and Harris, 2008). as the number of available positions for graduates in both public and private sectors is by far less compared to the graduates produced every year, it is logical to assume that some graduates will remain unemployed the very year they graduated. And since as time goes by, their chances of being employed reduces due to the increase population of unemployed graduates because new graduates are being produced, it is reasonable to say that many graduates will forever remain unemployed which represents cure fraction and refers in this work as unemployment fraction.

When cure fraction is present in a time-to-event data, the usual models such as AFT and CPH will not be appropriate in describing the data, this is because they are based on the assumption that every graduate will eventually be employed (susceptible), hence the need for a more appropriate survival model.

Cure survival models are used to estimate cure fraction as well as the survival function of the susceptible (unsure) sub-population (Cancho et al., 2019a; Othus, Barlogie, LeBlanc and Crowley, 2012; Scudilio et al., Sreedevi and Sankaran, 2021). The earlier works on cure fraction models were credited to Boag (1949) and Berkson and Gage (1952), where they developed what is called Mixture Cure Model (MCM). Based on their model, the population of time-to-event is assumed to be a mixture of cure and uncure subjects. Other cure models include Non-Mixture Cure Models (Yakovlev, Asselain, Bardou, Hoang and Tsodikov, 1993) and Defective models (Balka, Desmond and McNicholas, 2019). To our knowledge, no research utilizes cure models in fitting time-to-employment data.

The objectives of the paper are to estimate unemployment fraction, survival functions of the uncure/employment sub-group as well as effects of covariates on both unemployment probability and survival function of the employment sub-group from the time-to-employment data.

To achieve the objectives, parametric, non-parametric and semi-parametric techniques were proposed. The rest of the paper is organized as follows; section 2 provides details on the statistical procedures and models used in the study. Results of the analysis and discussions are presented in section 3, while section 4 gives conclusion.

METHODOLOGY

This section provides the methodology used to conduct the analysis. Brief description on the data used and important functions in survival data analysis will be given. Kaplan-Meier Estimator (Kaplan and Meier, 1952) would be employed to graph the employment curve and see visually the possibility of cure subjects in the data. Further analysis would follow based on the KM results.

Data

The data consists of information on graduates of statistics department (2007 – 2018), Kano University of Science and Technology, Kano, Nigeria. A total of 518 students graduated within the period. Data on 311 graduates were obtained and 273 were subsequently used because of completeness. Information obtained are from two sources; information on the number of graduates for each year, their names, registration number, gender and class of certificates were obtained from the University Academic Division Office. Information on age at graduation, whether employed or not and the date of employment if any are obtained from the individual graduates. Those not employed at the time of collecting data are referred to as censored observations. This data is referred in this work as Graduate Employment Status (GES). The information from the graduates was obtained from June – December 2021. The response variable (time-to-employment) is measured in months. In this study, three covariates are considered; Class of certificate/Grade, Age and Gender. Age is defined as the age of the graduate at the time of finishing mandatory National Youth Service Corps (NYSC) measured in years. NYSC is a one-year service to the nation that is mandatory for every graduate of not more than 30 years of age at the time of graduation. Gender is a dummy variable for female and male, while class of certificate refers to the grade of degree student graduated with, which has 4 categories; First class (FC), Second class

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

upper (SCU) division, Second class lower (SCL) division and Third-class (TC) degrees. Initially students can graduate with a Pass degree which is the least category but was later scrapped. As such the only 3 graduates with pass degrees were merged with those with TC degrees. Information was collected from the graduates using several methods with the help of trained assistants. We first obtained contacts of some graduates from their department of graduation and using those graduates to obtain other contacts. What's up group of various classes and Alumni were utilized and follow up on individual graduates when the need arises was also employed.

Survival Data Analysis

Let T be a non-negative random variable ($T \geq 0$) describing the length of time between graduation and getting employment. Therefore, $f(t), t \geq 0$ is the probability density of T, while $S(t) = P(T > t) = \int_t^{\infty} f(x)dx = 1 - F(t)$ is the survival function, which is the probability of a graduate to remain unemployed beyond time t, $F(t)$ is the cumulative distribution function.

The hazard function defined for $t > 0$, $h(t) = \frac{f(t)}{S(t)} = \lim_{\delta t \to 0} \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t}$. The hazard function here represents the probability that an unemployed graduate at time t experiences the event (employment) in the next period δt (Tibshirani, 1982). Here, hazard function is the conditional probability of getting employment in the next period given that a graduate is unemployed now. Therefore, it is graduate chance of getting employment at a given point in time as such gives an easiest way to describe important functions in survival analysis such as the Proportional Hazard Model.

The cumulative hazard function given by $H(t) = \int_0^t h(x)dx$ is related to the survival function as follows, $S(t) = exp(-H(t))$ or $H(t) = -logS(t)$. Survival models are built based on survival or hazard function and the two functions describe the survival experience of a population in a different way.

Kaplan – Meier

Kaplan and Meier (1952) introduced a non-parametric procedure of estimating baseline survival function even in the presence of censored observations. It is a step function curve, with each step representing the occurrence of the event of interest. Kaplan – Meier (KM) Estimator also known as Product Limit Estimator at time t is given by

$$\widehat{S_0}(t) = \prod_{i:t_i \leq t}^{n} \left(1 - \frac{d_i}{n_i}\right) \quad (1)$$

Where d_i represents the number of employed graduates and n_i, the number of graduates at risk of getting employment at time t_i. $S_0(\lambda)$ is taken to be the unemployment fraction of the population of study where λ is the largest observed event (Maller and Zhao, 1996). The presence of cure subjects is portrayed in the KM curve by a long stable plateau at the extreme right tail of the curve.

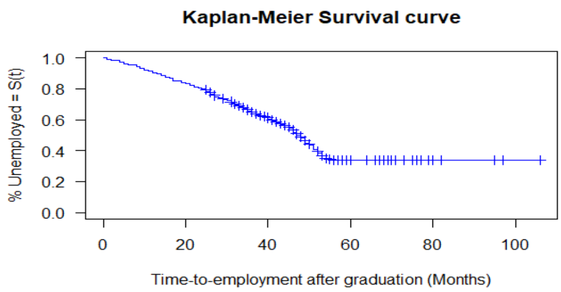

Fig 1. Kaplan – Meier for the entire data

The Kaplan Meier in fig 1 has shown the possibility of a cure fraction as such cure models are the appropriate models to use in analyzing the data. In this paper, Mixture Cure Model will be employed.

Log – Rank Test

Log – rank is a test used alongside KM to find out if survival is significantly different in a categorical variable. For example, if gender significantly influence survival.

Mixture Cure Model (MCM)

Mixture cure model consists of two types of subjects, the cure and the uncured subjects. Here, the cure subjects are the number of graduates that will forever remain unemployed here refers as unemployment fraction while unsure refers to potential graduate employees. MCM described by survival function can be given by the following unconditional population survival function

$$S(t) = P(T > t) = pS_u(t) + (1 - p)S_s(t) \qquad (2)$$

where $S_s(t)$ is the survival function of the potential employed graduates (susceptible)

$S_u(t)$ is the survival function of the graduate's unemployment fraction.

P is the probability of being unemployed

T is the Time – to – employment

t any specific time

Given Y as an unemployment indicator (Y = 0, for unemployed graduate and Y = 1 employed graduate), $S_u(t) = P(T > t/Y = 0) = 1$, is a degenerate survival function (Peng and Yu, 2021), because it is certain that all cured subjects will survive beyond any time t. Therefore,

$$S(t) = p + (1 - p)S_s(t) \qquad (3)$$

This shows that, the MCM is a two-parts model, the part describing the probability of being unemployed called Incidence, and the part describing the distribution of the survival times of the potential employees called Latency. Therefore, MCM strategy involves harmonizing the two models representing the two parts. It is important to note that, the probability of being unemployed (Incidence part) and the survival function of the potential graduate employees (Latency part) can be influenced by similar or different covariates.

The incidence part is always described by Bernoulli distribution, since the random variable Y is binary. Logit function will be used to link the effect of covariates (z) to the unemployment probability (p). According to Farewell (1982), $logit[p(z)] = Z'\gamma$ where $Z'\gamma$ are linear predictors and γ is the vector of the coefficients of z, therefore,

$$p(z) = \left(\frac{\exp(Z'\gamma)}{1+\exp(Z'\gamma)}\right) \qquad (4)$$

In this paper, the latency part will be described using both CPH and Accelerated failure time assumptions. The resulting Mixture Cure Models would be called CPH MCM and AFT MCM. A distribution that is found to describe the time-to-employment data best would be the baseline distribution in the AFT MCM.

Models of the latency part of the MCM/Survival Regression Models

Here, two models will be considered; Accelerated Failure Time (AFT) Model and Cox Proportional Hazard Model (CPHM)

Accelerated Failure Time (AFT) Model

In AFT model, a covariate effect either accelerates or decelerates the time-to-failure by some constants. Here, a distribution is selected depending on how best it fits the time-to-event data. AFT allows measuring the direct effect of the covariates on the survival time. This makes interpretation of effect of covariate on the mean survival time easier (Gelfand, Mackinnon, DeRubeis and Baraldi, 2016). Table 1 gives the fit of five popular survival models to the employment data. Based on Log-Likelihood (LL) and AIC criterion, Weibull distribution fits the data best as such would be used in building the AFT MCM.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

An AFT model as a function of explanatory variables X's can be given by the following equation

$$\ln(T) = \mu + \beta' x + \sigma\epsilon \quad (5)$$

Where μ is the intercept, $\beta' = (\beta_1, \cdots, \beta_p)$ a vector of regression coefficients, σ a scale parameter and ϵ, an error term following a particular distribution. The choosing distribution of an error term give rise to a specific AFT regression model.

Table 1 fit of some models using log-likelihood and AIC

Model	Log-likelihood	AIC
Exponential	-747.63	1503.26
Weibull	-738.70	1477.40
Lognormal	-745.83	1501.66
Logistic	-764.77	1539.54
Log-logistic	-739.14	1488.28

Weibull Accelerated Failure Time Mixture Cure Model (Weibull AFT MCM)

Supposing X is a vector of covariates that influence the survival time of the employed graduates (uncured sub-population) and β is the vector of the coefficients in X, then based on Cox and Oakes (1984), the latency sub model ($S_s(t/x)$) is given as the function of the survival function of a baseline distribution ($S_{s0}(t)$) and linear predictors ($\beta' x$), that is

$$S_s(t/x) = S_{s0}(te^{-x'\beta}) \quad (6)$$

According to Peng and Yu (2021),

If $Log(T/Y = 1) = x'\beta + \sigma\epsilon$ (7)

With σ as a scale parameter and ϵ an error term satisfying $P(e^{\sigma\epsilon} > t) = S_{s0}(t)$, then $T/Y = 1$ will follow the model (6).

With Weibull as the baseline distribution describing time-to-employment data, the error term (ϵ) is assumed to follow an extreme value distribution, as such the resulting AFT will be Weibull. Therefore, the survival function of the employment sub population following Weibull distribution given covariates Xs is given by

$$S_s(t|x) = \exp[-(te^{-x'\beta})^{\frac{1}{\sigma}}] \quad (8)$$

Therefore, our corresponding MCM is a Weibull AFT MCM given by

$$S(t/x) = p(x) + (1 - p(x))\exp[-(te^{-x'\beta})^{\frac{1}{\sigma}}] \quad (9)$$

Here, the unemployment fraction, $p(x) = \lim_{t\to\infty} S(t/x)$

Cox Proportional Hazard Model (CPHM)

Semi-parametric survival regression model known as Cox Proportional Hazard Model (Cox, 1972), is the popular survival regression model used to ascertain the influence of covariates on survival, modeled through hazard function. Here, effects of covariates are measured on the hazard instead of survival as in AFT model.

Based on this model the hazard of getting employment of a graduate at time t with a given set of time-independent explanatory variables x, is given by

$$h(t|x) = h_0(t)\exp(\beta_1 x_1, \beta_2 x_2, \dots, \beta_p x_p)$$

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

where $h_0(t)$ is called the baseline hazard function and β, the regression coefficient.

CPHM as semi-parametric model makes no assumptions about the shape of the baseline hazard function. But irrespective of its shape, it is the same for everyone. The choice of CPHM is appropriate or even better if proportional assumption is satisfied, or if we are not confident about the shape of the baseline hazard. Proportional hazards assumption will be tested using Schoenfeld test procedure and subsequently select covariates to be used in the semi-parametric CPH mixture cure model.

Cox Proportional Hazard Mixture Cure Model (CPH MCM)

if X is a vector of covariates that influence the survival time of the employed graduates (susceptible sub-population) and β is the vector of the coefficients in X, then based on the proportional hazard assumption, then the latency part of the MCM is given as

$S_s(t/x) = S_{s0}(t)^{\exp(x'\beta)}$ while the hazard function given by

$h_s(t/x) = h_{s0}(t)^{\exp(\beta' x)}$ and the corresponding CPH MCM is given by

$$S(t/x) = p(x) + (1 - p(x))S_{s0}(t)^{\exp(x'\beta)} \qquad (10)$$

Therefore, the cure fraction ($p(x)$) is obtain as

$$\lim_{t\to\infty} S(t/x)$$

Note that, the baseline survival/hazard function is unspecified as in the usual CPH model.

Model Assessment

Models in this work will be compared based on either Loglikelihood (LL) or Akaike information criterion. Akaike (1974) proposed criterion to be used in selecting best model among both nested and non-nested models. The AIC used in this work is given by

$$AIC = -2\log(L) + 2(k + c)$$

Here, L is the likelihood value, k, number of covariates in the model and c, the number of model-specific distributional parameters (Bradburn, Clark, Love and Altman, 2003). AIC is a measure of error and as such a model with lower AIC is preferred than the one with higher AIC. Unlike AIC, higher value of Loglikelihood indicates better model. All our models will be analyzed using functions in the R statistical software version 4.0.4.

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 has shown the distribution of our categorical demographic variables. Majority of the graduates aremales constituting about 87% while females are about 13%. With respect to gender, the employment percentage of males is higher while that of second-class upper is higher with respect to grade. Most of the students graduated with SCL degree.

Table 2 Demographic variables

covariates	Sample Size (n)	Employed (%)	Unemployed (%)
Gender			
Female	35	16 (45.7)	19 (54.3)

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

Male	238	125 (52.5)	113 (47.5)
Grade			
First Class	6	4 (66.7)	2 (33.3)
Second Class Upper	52	35 (67.3)	17 (32.7)
Second Class Lower	142	63 (44.4)	79 (55.6)
TC	73	39 (53.4)	34 (46.6)

Kaplan Meier

The Kaplan Meier in fig 1 has shown the possibility of an unemployment fraction. The estimated fraction is 33.8% with a median survival time of 47 months as shown in table 3. KM was also constructed for gender (fig 2) and grade (fig 3). As shown in table 3, the KM estimates of unemployment fraction is 47.7% for females and 32.9% for males. Based on the log-rank test, gender is not statistically significant (P = 0.8) in influencing time-to-employment. The KM estimate of unemployment fraction with respect to grade has shown that, SCL graduates has the highest unemployment proportion and SCU has the least. Grade is found to be significant in influencing time-to-employment based on the log-rank test (P = 0.0003), that means employment experience of a graduate is significantly determine by his/her grade of certificate. Table 3 has also given the result of Schoenfeld tests where both Gender and Grade were found to satisfy PH assumption at 5% level with P = 0.067 and P = 0.46 respectively as such can be included in our models. The only continuous variable in the research, Age, also satisfied the PH assumption (P = 0.41). Most researchers would not include variables that are not significant in either AFT or CPH Models (eg Gender), but we wish to include it here since we are dealing with cure models, this is because the nature or strength of its relationship with survival may change in the employment sub-population when cure is removed. In addition, we would like to know as part of our objectives whether a covariate also influences unemployment probability. Therefore, we wish to build the AFT and CPH Mixture Cure Models with each of the covariate as the sole explanatory variable. Since in literature researchers considers only non-cure survival regression models (CPH or AFT) to analyze time-to-employment data, we would also construct CPHM and AFTM comparison withtheirMCMcounterpart.ForcomparisonwiththeirMCMcounterpart.

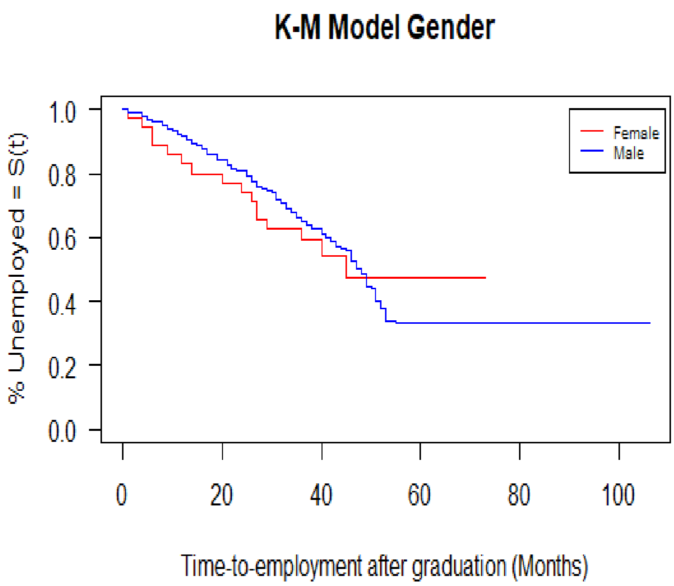

Fig 2. Kaplan – Meier Survival Curves according to Gender

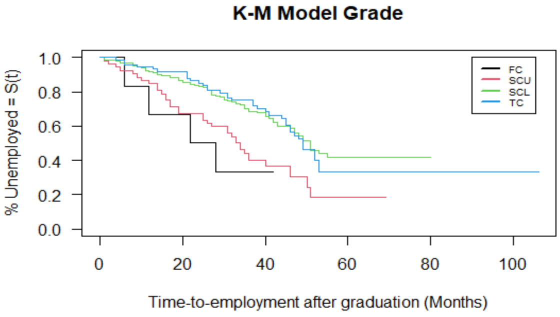

Fig 3. Kaplan – Meier Survival Curves according to Grade

Table 3 KM estimate of Unemployment fraction, Log-rank and Schoenfeld tests

Covariates	Cure Fraction (%)	Median Survival Time	Log-rank Test (P-value)	Remark	PH Assumption (P-value)	Remarks
Entire Data	33.8	47	-	-	-	-
Gender						
Female	47.7	45	P = 0.8	Not Sig	0.067	Satisfied
Male	32.9	48				
Grade						
FC	33.3	25				
SCU	18.2	33	P = 0.0003	Sig	0.46	Satisfied
SCL	41.6	51				
TC	33.3	49				
Age	-	-	-	-	0.41	Satisfied

CPH MCM versus Weibull AFT MCM

As shown in table 4, CPH MCM gives a closer estimate of unemployment fraction to the empirical estimate by Kaplan-Meier as well as lower AIC compared to Weibull AFT MCM. None of the covariates in the two models was found to significantly influence unemployment fraction.

Table 5 compares CPH MCM and Weibull AFT MCM. All the regression parameters (coefficient) in CPH MCM are < 0. This means that, each of the reference group have higher hazard of getting employment than the group/s of interest. For example, females as the reference group have higher hazard as such get employment faster than males. The HR of 2.1778 implies that at any given point in time, the hazard of getting employment of females is 2.18 times that of males, or at any given point in time, females have about 118% more likely to get employment than males. More importantly the difference in survival with respect to gender is significant in the CPH MCM. Age was also significant in the CPH MCM, but grade was not significant.

With respect to Weibull AFT MCM, all the regression parameters of the covariates are greater than zero. This means that, all the groups of interest have higher survival probability than the reference group. That is, they stay more in the labor market compared to the reference group. For example, at any given point in time, males have 1.58 times more likely to stay in labor market than females or males have to wait 1.58 times to get employment than females. But, none of the covariates (Gender, Grade and Age) is significant in influencing time-to-employment in the latency part of the Weibull AFT MCM. This can partly be attributed with the fact

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

that, Weibull AFT MCM could not effectively estimate cure fraction like CPH MCM, otherwise would have shown the significant influence of Age and Gender on time-to-employment of the susceptible sub population like was revealed by CPH MCM. None of the covariates in each of the models significantly influence the probability of unemployment in the incidence sub-model.

Table 4 Cure Fraction Estimates from Cure models

Model	Kaplan-Meier	CPH MCM		Weibull AFT MCM	
Covariate	**Cure Fraction**	**Cure Fraction**	**P-Value**	**AIC**	**Cure Fraction**
Entire Data	33.8	28.6	-	1427.83	24.8
Gender					
Female	47.7	47.5			45.5
Male	32.9	32.9	0.2215	1483.29	24.1
Grade					
FC	33.3	31.6			31.5
SCU	18.2	16.9	0.9433		8.64
SCL	41.6	41.5	0.9859		29.4
TC	33.3	33.9			24.9
Age	-	-	0.5874	1425.21	-

Table 5 Analysis of CPH MCM and Weibull AFT MCM

Model	Sub-Model	Covariates	Coefficients	HR/TR	SE(Coeff)	P-Value
CPH MCM	**Gender**	RG = Female, Male	-0.7783	2.1778	0.3536	0.0277*
	Latency	Grade RG = FC				
		SCU	-0.9028	2.4665	1.8425	0.6242
		SCL	-1.3132	3.7181	1.8161	0.4696
		TC	-1.5354	4.6432	1.8317	0.4019
	Age	Continuous	-0.1032	1.1088	0.0481	0.0317*
	Incidence					
	Gender	Male		0.1738		
	Grade	SCU		0.9017		
		SCL		0.9433		
		TC		0.9859		
	Age	Continuous		0.5874		
WAFT MCM	**Gender**	RG = Female, Male	0.4558	1.58	0.2824	0.1065
	Latency	Grade RG = FC				

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VIII Issue XII December 2024

		SCU	0.6238	1.87	1.3112	0.6343
		SCL	0.8983	2.46	1.2994	0.4894
		TC	0.9165	2.5	1.3032	0.4819
	Age	Continuous	0.0335	1.034	0.0285	0.2388
	Incidence					
	Gender	Male		0.2215		
	Grade	SCU		0.7912		
		SCL		0.9832		
		TC		0.9449		
	Age	Continuous		0.482		

Note: For easier interpretation of HR, we swap reference group with group of interest (exp$^{-(\text{coefficient/loghazard})}$

CPHM versus CPH MCM

The regression parameter (coefficient) in both CPHM and CPH MCM for Gender is less than zero (Table 6). This means that, females as the reference group have higher hazard as such get employment faster than males. While females and males have almost similar hazard of getting employment in the CPHM (1.0574), the difference is large in CPH MCM (2.1778). More importantly the difference in survival with respect to gender is only significant in the CPH MCM. With respect to Grade all the regression parameters are less than zero in both CPHM and CPH MCM which implies higher hazard/chances of getting employment for FC graduates (reference group) compared to all other graduates. But the difference is only significant between FC and SCL graduates in the CPHM. Also like in gender, the difference in hazard of getting employment between FC and each of the other graduates increases when cure was removed as shown in table 6. For example, at any given point in time, FC graduates have over 170% more likely to get employment than the TC graduates in the CPHM, but the percentage difference is over 360 in the CPH MCM.

Age is only significant in influencing employment in the CPH MCM with younger graduates having higher hazard of employment compared to older graduates. Here also, the hazard ratio is higher in the mixture model. We can see that, while a graduate who is one year younger has 5% more hazard of getting employment compared to a graduate who is one year older in non-cure model (CPH), the percentage more than doubles (11%) in the cure model (CPH MCM).

Table 6 Analysis of CPH Model and CPH MCM

Model	Sub-Model	Covariates	Coefficients	HR/TR	SE(Coeff)	P-Value	Coefficients (MCM)	HR (MCM)	SE(Coeff) (MCM)	P-Value (MCM)
CPHM	**Gender**	RG = Female, Male	-0.0558	1.0574	0.2658	0.834	-0.7783	2.1778	0.3536	0.0277*
Latency	**Grade**	RG = FC								
		SCU	-0.2977	1.347	0.5339	0.5771	-0.9028	2.4665	1.8425	0.6242
		SCL	-1.0689	2.912	0.5213	0.0404*	-1.3132	3.7181	1.8161	0.4696

		TC	-0.9972	2.711	0.5317	0.0607	-1.5354	4.6432	1.8317	0.4019
Age (Continuous)			-0.0528	1.054	0.0458	0.2489	-0.1032	1.1088	0.0481	0.0317*
Incidence	**Gender**	Male		0.1738						
	Grade	SCU		0.9017						
		SCL		0.9433						
		TC		0.9859						
Age				0.5874						

Note: For easier interpretation of HR, we swap reference group with group of interest (exp$^{-(\text{coefficient}/\text{loghazard})}$)

CONCLUSION

The analysis revealed that the unemployment proportion tends to inflate the time-to-employment function; therefore, when it is removed, the true influence of variables on time-to-employment functions emerges. The status may even change from insignificant to significant or vice versa, as we have seen in the CPH MCM with gender and age going from insignificant to significant and grade changing from significant to insignificant, respectively. This demonstrates the advantage of cure models in revealing the true effect of variables on survival. The findings may aid the government and other stakeholders in employment planning policies to avoid a high percentage of unemployment, which leads to poverty, and poverty leads to insecurity. The rising rate of unemployment, which affects all nations, may be linked to the current global degree of insecurity.

ACKNOWLEDGEMENT

Authors wish to acknowledge Nigerian Government through Tertiary Education Trust Fund for sponsorship of the first author and University Technology Malaysia for providing enabling environment as well as Kano University of Science and Technology for providing the data for the research.

CONFLICT OF INTEREST

No conflict of interest reported by the authors

REFERENCES

1. Manuel, S. (2007). The Transition from Higher Education to Employment in Europe: The Analysis of the Time to obtain the First Job. Higher Education, 54, 333–360.
2. Lim, H. (2008). Feasibility of Early Identification of Low Employability Graduates in Malaysia. Malaysian journal of economic studies, 45(2), 95 – 112.
3. Barros, C. P., Guironnet, J. P., & Peypoch, N. (2010). How to quickly get a job? The transition from higher education to French labor market by a survival model. Applied Economics, 43(4), 439-448.
4. Rosna, A., Lim, H. E., Bidin, Y., Tengku, F. T. A., Ainol, M. Z., Haniza, Y. & Omar, O. (2015). Estimating A Prediction Model for The Early Identification of Low Employability Graduates in Malaysia. The Singapore Economic Review, 60(4), 1550027 © World Scientific Publishing Company.
5. Ezeani, E. (2018). Barriers to graduate employment and entrepreneurship in Nigeria. Journal of Entrepreneurship in Emerging Economies, 10(3), 428 – 446.
6. Soon, J., Lee, A. S., Lim, H., Idris, I. & Eng, W. Y. (2019). Cubicles or corner offices? Effects of academic performance on university graduates' employment likelihood and salary. Studies in Higher Education,

7. Ayaneh, M. G., Dessie, A. A. & Ayele, A. W. (2020). Survival Models for the Analysis of Waiting Time to First Employment of New Graduates: A Case of 2018 Debre Mar University Graduates, Northwest Ethiopia. Education Research International, Article ID8877504.
8. Cox, D. R. (1972). Regression models and life tables. Journal of the Royal Statistical Society Series B. 34(2): 187-220.
9. Federal Government of Nigeria (2017), “Key statistics labor market, unemployment” National Bureau of Statistics, available at: (accessed 26 September 2021).
10. Lim, H. E., Rich, J. & Harris, M. (2008). Employment outcomes of graduates: the case of University Utara, Malaysia. Asian Economic Journal, 22(3), 321 – 341.
11. Othus, M., Barlogie, B., LeBlanc, M. L., & Crowley, J. J. (2012). Cure models as a useful statistical tool for analyzing survival. Clinical Cancer Research, 18(14), 3731-3736.
12. Scudilio, J., Calsavara, V. F., Rocha, R., Louzada, F., Tomazella, V. & Rodrigues, A. S. 2018). Defective models induced by gamma frailty term for survival data with cured fraction. J. Appl. Stat. 46: 484–507.
13. Cancho, V. G., Macera, M. A. C., Adriano K., Suzuki, A. K., Louzada, F. & Zavaleta, K. E. C. (2019a). A new long-term survival model with dispersion induced by discrete frailty. Lifetime Data Analysis.
14. Sreedevi, E. P. & Sankaran, P. G. (2021). Statistical methods for estimating cure fraction of COVID-19 patients in India. Model Assisted Statistics and Applications 16: 59 – 64. 15. Boag, J.W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of Royal Statistical Society-Series B, 11(1), 15-53.
15. Berkson, A. J & Gage, R. P. (1952). Survival Curve for Cancer Patients Following Treatment. Journal of the American Statistical Association, 47, 501 – 515.
16. Yakovlev, A.Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, R., Rochefediere, A. & Tsodikov, A. D. (1993). A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer as cited in Balka, J., Desmond, A. F. & McNicholas, P. D. 2009. Review and implementation of cure models based on first hitting times for Wiener processes. Lifetime Data Anal, 15, 147–176.
17. Balka, J., Desmond, A. F. & McNicholas, P. D. (2009). Review and implementation of cure models based on first hitting times for Wiener processes. Lifetime Data Anal, 15, 147– 176.
18. Kaplan, E. L. & P. Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53 (282), 457–481.
19. Tibshirani, R. (1982). A Plain Man’s Guide to the Proportional Hazards Model. Clinical Investigative Medicine, 5, 63 – 68.
20. Maller, R., & X, Zhao. (1996). Survival Analysis with Long-Term Survivors. Wiley, New rk.
21. Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-rvivors. Biometrics, 38, 1041–1046.
22. Gelfand, L. A., MacKinnon, D. P., DeRubeis, R. J. & A. N. Baraldi, A. N. (2016). Mediation Analysis with survival outcomes: accelerated failure time vs. proportional hazards models. Frontiers in Psychology, 7, 423.
23. Cox D. R. & Oakes D. (1984). Analysis of Survival Data. New York, NY Chapman and Hall Ltd.
24. Peng, Y. & Yu, B. (2021). Cure Models: Methods, Applications and Implementation. Chapman & Hall/CRC Biostatistics Series, CRC Press, Taylor and Francia Group.
25. Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans Automatic Control, 19, 716 – 723.
26. Bradburn, M.J., Clark, T. J., Love, S. B. & Altman, D. G. (2003). Survival analysis part III: Multivariate data analysis choosing a model and assessing its adequacy and fit. British Journal of Cancer, 89: 605 – 611.