International Journal of Applied Business and Management Sciences
Vol. 3, No. 2, 2022, pp. 267-288
https://DOI: 10.47509/IJABMS.2022.v03i02.07

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

A STATISTICAL ANALYSIS OF TIME TO A CLAIM: A CASE OF ZIMBABWE'S HEALTH INSURANCE CLIENTELE

Desmond Mwembe[1], Brian Jones[2], Wilbert Chagwiza[3] and Solwayo Ngwenya[4]

[1]Lecturer, Department of Statistics and Operations Research, National University of Science and Technology, Bulawayo, Zimbabwe. E-mail: desmond.mwembe@nust.ac.zw
[2]Professor Emeritus, Department of Statistics and Operations Research, National University of Science and Technology, Bulawayo, Zimbabwe. E-mail: amabooksbyo@gmail.com
[3]Manager: Data Scientist, Standard Bank of South Africa Head Office, Fraud Risk Management, 30 Rosebank, Johannesburg. E-mail: wilbertchagwiza@gmail.com
[4]Professor of Obstetrics & Gynaecologist, Head of Department of Obstetrics & Gynaecology Clinical Director, Mpilo Central Hospital, Vera Road, Mzilikazi. E-mail: drsolngwe@yahoo.co.uk

Article History
Received : 02 November 2022; Revised : 20 November 2022; Accepted : 05 December 2022; Published : 29 December 2022

Abstract: The aim of this study was to find factors influencing time to a claim among the insured population of Zimbabwe. The factors under consideration were age, sex, medical aid scheme and amount claimed from the medical aid society. These were chosen on the basis of the variables that were collected at hospitals where claims were made. Claims data between 2013 and 2016 from the health insurance companies were used. Data were collected from government hospitals and were then cleaned and analysed using R-package and SPSS. Cox regression model from the survival analysis was proposed to find the influential factors in the study. It was established that without interactions, age, sex, and scheme type were influential in shortening time to a claim. After introduction of interactions, scheme type and amount claimed were found to be highly influential in shortening time to a claim. It was concluded that clients with medium premium schemes had a shorter time to a claim irrespective of their gender, medical aid society, age, and amount claimed. The study recommended that medical aid societies should know the expected time to a claim in each medical aid scheme and likely range of amount to be claimed by their clients in order to optimize benefits packages to cover clients' medical expenses, yet remain within sustainable operating limits.

Keywords: Health, insurance, Cox model, time to a claim, Zimbabwe

To cite this paper:
Desmond Mwembe, Brian Jones, Wilbert Chagwiza and Solwayo Ngwenya (2022). A statistical analysis of time to a claim: A case of Zimbabwe's health insurance clientele. *International Journal of Applied Business and Management Sciences*. 3(2), 267-288. *https://DOI: 10.47509/IJABMS.2022.v03i01.07*

INTRODUCTION AND BACKGROUND

There are many insurance companies in Zimbabwe, there are some in-house medical aid companies for parastatals and others are open societies (AFHoZ, 2008). These companies have developed schemes that clients can choose to be members. Selection of a scheme depends on many characteristics of a client. These characteristics include employment status of the client, age of the client, source of income, health condition and many other. Medical aid companies have different schemes that are related but under different names. In this research, three categories of medical schemes are developed. These are Low premium schemes, medium premium schemes and high premium schemes (Mwembe *et al*, 2020).

This research focuses on analysing time to a claim by clients with different characteristics. This research analyses health claims frequencies data from insurance companies, as from 2013 to 2016, during the dollarization era. During this period, the inflation was at a minimum as compared during the period when the Zimbabwean dollar is in use. Data during the period of the Zimbabwean dollar give inflated estimates due to inflation, this distorts distribution, and estimation of future claims will be highly be affected.

Survival of insurance companies depends on the frequency of claiming by enrollees. The higher the frequency, the lower the profits realised in insurance business. As Rendek *et al* (2014) says, an insurance company is said to be performing its mandated responsibilities if it settles or responds to claims made by its clients. If there is no response made by insurance companies, then it ceases to be realistic and enrollees will not appreciate the worthiness of their subscriptions. Failure to settle claims made by enrollees swiftly results in increase in out-of-pocket payments (Dalal *et al*, 2014). As a pool of risks, premiums from different clients act as a cushion whenever a hostile circumstance befalls an enrollee.

The attainment of the notion of being a pool of risks, time to a claim by enrollees needs understanding to a deeper extent in health insurance companies. This improves planning by health insurance companies in the event of a claim made by a client. If the distribution of claiming by a set of clients with almost similar characteristics is established, prediction of a claim with given severity will be possible. Modelling claims will make it possible for health insurance companies to estimate the number of claims clients are likely to make and hence plan ahead of time to settle claims that are likely to be made. This will drastically reduce churn of clients from one medical aid company to another or churn from one medical scheme to another within a medical aid company due to disgruntlements.

PROBLEM STATEMENT AND SIGNIFICANCE OF STUDY

Many studies in health insurance are mostly concerned with the characteristics of clients influencing choice of a medical scheme (Qian *et al* 2010, Sello *et al*, 2012). Other studies are on adverse scheme selection, pricing models and coverage of individuals by schemes in different setups. Insurance companies are also faced with problem of frequency of medical claims. The length of time between claims by a client is crucial. This has an effect on the medical aid companies' financial reserves. The shorter the time between claims, the higher the cumulative amount claimed by an individual and this lowers financial reserves of a medical society. Studies on the frequency of claims have not been done in Zimbabwe. Knowledge on how long clients take to make a claim does not exist among Zimbabwean health insurance companies. This has repercussions of being overwhelmed with claims and failing to settle some of the claims. Findings of this study can be used by health insurance policies in forecasting future number of claims. This reduces chances of being overwhelmed by claims. Even if there are more claims than expected, the degree of being overwhelmed can not be compared to when there was no estimation made beforehand. A health insurance provider therefore can use this knowledge to plan well ahead of time and increase customer satisfaction and retain as many clients as possible. It also helps health insurance providers optimise benefits packages to cover clients' medical expenses, yet remain within sustainable operating limits.

AIM

To model average time between claims by clients with different characteristics in Zimbabwe

LITERATURE REVIEW

There are many forms of data in many different statistical disciplines. Time to a claim is an example of time to an event data and hence is classified as survival data. This data is analysed using survival methods that include distributions like the Cox regression (Cox, 1972). In this research, use of Cox regression model is to assess the influence of several characteristics of medical insurance enrollees on their frequency to a claim.

Gustafsson (2009) did prediction of policy churns in non-life insurance. This was done in motor industry and this prediction lacks in life insurance companies, particularly in Zimbabwe. Insurance data is highly skewed, Cox regression model in this research will be used because, the data is skewed and there is censoring. Clients will have made a claim or not during the study period. Mashasha *et al* (2022) analysed claims of health insurance enrollees from a health insurance company in Zimbabwe. The researcher used time series approach of which, given the skewed nature of the insurance data, time series does not precisely model insurance data. Xie (2016) used data mining methodology to predict time of hospitalization of insured clients. These are all efforts to estimate the amount of claims that are likely to result from a given individual. The research by other scholars do not incorporate the individual characteristics of these insurance companies or schemes. This deficiency does not estimate exactly the number of claims an individual is likely to make. Cox regression model will incorporate the characteristics of individual clients and the skewedness of the data that other methodologies do not incorporate.

Patricia (2014) examined data from a National Insurance Commission and introduced methodologies of survival analysis to model the average time of how motor insurance claims are handled and the variables that are affected in the automobile insurance industry over a one year period. The explanatory variables in this study were age, gender, marital status, the type of vehicle involved in the accident, type of insurance policy bought, and the nature of the claim. Time was used as the response variable in this study, it was defined as the length of having an insurance policy until a loss occurs and when it was paid. To facilitate the analysis of the data the variable age was grouped into four categories, that is, 21-29years, 30-45years, 46-59years and 60years and above.

The nature of the claim as to whether a third party policyholder or a comprehensive policyholder was involved in an accident was considered. Patricia (2014) applied Cox regression analysis to determine the type of insurance and variables that affect the time it takes for a claim to be settled. The results showed that the type of vehicle and type of insurance policy were the only two significant variables in the study. After accounting for type of policies, and type of vehicle, there were no statistically significant associations between age, gender, marital status and nature of claim. This was not to say that these risk factors were not associated with claims; their lack of significance was likely due to confounding (interrelationships among the risk factors considered.

Zhang (2010) worked on claims in insurance companies with bankruptcy and acquisition being events of interest, which were correlated and censored.

Firstly, they assessed the results using the Cox proportional hazards model and then applied the copula function. Gascon *et al* (2016) in America looked at fiiting probability distribution to claims data. The focus was basically on health care costs but did not incorporate individual characteristics of clients. This still has a deficit on the characteristics of clients. As pointed out in the research, if clinical conditions are included, the prediction will be more precise and this necessitates a research that will take into account other factors of particular enrolees into account. Cekici *et al* (2018) and Konrad *et al* (2019) indicate that there are challenges associated with use of insurance data and no one-size-fit-all method can be used.

All the reviewed scholars that made use of the Cox regression model used maximum partial likelihood estimation to estimate the parameters and scaled Schoenfeld residuals to test for the proportional hazard assumption was used in this study. Zhang (2010) did not take into account the other two assumptions of the Cox regression model which are scrutinizing leading observations (or outliers) and detecting non-linearity in association between the log hazard and the independent variables. This research contributes in that it will scrutinise the individual characteristics of clients in influencing time to a claim.

METHOD

Data collection and cleaning

Secondary data was used from the health insurance companies. The study is based on the claims data of all participants aged 21 or older. 1015 subjects were selected from the claims database.

Claims data were collected and generated from the medical aid societies (MAS) in Zimbabwe. Medical aid societies (MAS) were classified into two, that is:

1. Private
2. Government

Information on age, sex, medical aid society, scheme type (low premium scheme, medium premium scheme, high premium scheme), date of claim and amount claimed was extracted from the documents of the insured clients who made claims during the dollarization era. Four consecutive years of data, from 2013 to 2016, were provided for analysis.

Variables in the Study

The study comprises the response and explanatory variables. The dependent variable or response is the average time between claims. Predictors or explanatory variables, which are called covariates, are those whose effects on average time between claims are assessed. The predictor (covariate) variables which are assumed to be influencing time to a claim among the clients are:

- Age,
- Gender,
- Medical aid society,
- Scheme type,
- Amount claimed.

Model and Estimating Model Parameters

Given the data characteristics of average time between claims for a dependent variable, some clients might have claimed more than once in the period 2013-2016. Dates when claims were made were recorded. Independent variables were both numerical and categorical in nature and could be grouped into categories such as male or female. Due to these qualities of both the dependent and independent variables, a Cox regression model is suggested to best explain the data.

Maximum likelihood estimation was used in estimation of model parameters. The Cox model enables the researcher to get hazard ratio. This function is always nonnegative and so is the confidence interval. The hazard rate is the probability that an event not occurring in the interval [0, t], it will occur in the interval [0, t+1] divided by the width of the interval under consideration. This interval normally made small so that the hazard rate consequently signifies an instantaneous rate. In fitting such a model, the constant assumption of the proportional hazards regression is invoked.

The Cox regression model is mathematically written as follows:

$$h_i(t) = h_0(t)\exp(b_1x_1 + b_2x_2 + \cdots + b_nx_n)$$

where,

$h_i(t)$ is the expected hazard of the i^{th} individual at time t.

The covariates of $x_1, x_2, \ldots, x_n$ are explanatory variables.

$h_0(t)$ is the baseline hazard function that denote the hazard of the i^{th} individual when all of the explanatory variables $x_1, x_2, \ldots, x_n$ are equal to zero.

The exponentiated coefficients (exp (b_i)), also known as hazard ratios, give the effect size of covariates. If the hazard ratio is above 1.0 and the confidence interval is entirely above 1.0, then exposure to the predictor increases the rate of time to a claim. If the hazard ratio is below 1.0 and the confidence interval is entirely below 1.0, then exposure to the predictor decreases the rate of claims. The ratio of the hazard at time t to the baseline hazard is given by:

$$\frac{h(t)}{h_0(t)} = \exp(b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)$$

If the model has positive coefficients for the explanatory variables, the hazards will increase (average time to a claim decreases) and the hazards decrease (average time to a claim increases) when the regression coefficient is negative. The p-value for likelihood ratio, Wald and score for log rank tests for overall significance of the model will be specified. For a large enough number of observations (N), the tests will have the same p-value.

Testing model coefficients

The p-value of less than 0.05 was used to test for the significance of model parameters. If the p-value is significant, $p<0.05$, it implies that the variable corresponding to that p-value will be included in the model and has an influence in predicting time to a claim. If the variable has a p-value greater than 0.05, $p>0.05$, the variable will be dropped out of the model. This indicates that the variable is not significant and has no influence in determining time to a claim, holding other explanatory variables in the model constant.

Test for proportional hazard assumption

According to Xue *et al* (2013), the proportional hazards model tries to simplify the constant multiplicative effect of each explanatory variable in the hazards function, over time. There is need to check whather the assumption was violated or not to obtain realistic results. The proportional hazard assumption can be checked using statistical tests and graphical diagnostics based on the scaled Schoenfeld residuals.

Test for outliers

Deviance residuals will be used to test for outliers in the data. They are used to establish improperly recorded observations. Jennings (1986) stipulates that deviance residual has a more balanced distribution about zero and for observation *i,* it is defined as a function of the Martingale residual (M_i):

$$d_i = sign(M_i)[-2(M_i + \delta_i \log(\delta_i - M_i))]^{\frac{1}{2}}$$

Points plotting outside or on the extremes of the deviance residuals were regarded to be outside the range and were termed outliers.

Testing non-linearity

It is highly assumed that continuous explanatory variables have a linear relationship. In fact, this should not end at assumption level but rather ideal to check whether the data does not violate this assumption. The commonly used methodology to check for non-linearity or functional form of covariate is by plotting martingale residuals against continuous covariates.

Consideration for non-linearity is for continuous variables instead of categorical variables. Configurations in plots of martingale and partial residuals versus a continuous variable may suggest that a covariate does not properly fit. Only continuous variables were considered in this research.

RESULTS

Descriptive Analysis

Table 1 and 2 describes the explanatory variables used in the estimation of the Cox regression model. The explanatory variables include age, gender(female:0, male:1), medical aid scheme (MAS) (government:0, private:1), scheme type (low premium scheme:1, medium premium scheme:2, high premium scheme:3) and amount claimed.

The participants' age ranged from 21 to 82 years old, with mean 54.05, median 53 and mode 51.

Table 1: Descriptive statistics for age and amount claimed

Variable	*Mean*	*Median*	*Mode*	*Standard deviation*	*Minimum*	*Maximum*	*Skewness*	*Kurtosis*
Age	54.05	53	51	3.835	21	82	0.9372	-0.1760
Amount claimed (US$)	147.87	94.00	20.00	170.398	17.48	1241.00	3.1827	13.7735
Time between claims (days)	64	51	-	56.478	0	334	1.9406	4.8330

The dataset is moderately and positively skewed (0.937). This is highly expected of the claims data and this implies that it is not normally distributed. Kurtosis of the dataset -0.1760 shows that it is light-tailed and has a platykurtic distribution hence not normally distributed. The data shows that there is high standard deviation in age, amount claimed and time between claims. This is evidence enough to show potential outliers in the data set.

Table 2: Gender composition in each medical Aid scheme

		Medical Aid Scheme		*Total*
		Government scheme	*Private scheme*	
Gender	Male	340	250	590(58.1%)
	Female	180	245	425(41.9%)
Total		520(51.2%)	495(48.8%)	1015

Table 3: Chi-Square Tests for Gender Compared to Medical aid Scheme

	Value	*df*	*Asymp. Sig. (2-sided)*	*Exact Sig. (2-sided)*	*Exact Sig. (1-sided)*
Pearson Chi-Square	23.068[a]	1	.000		
Continuity Correction[b]	22.461	1	.000		
Likelihood Ratio	23.147	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	23.045	1	.000		
N of Valid Cases[b]	1015				
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 207.27.					
b. Computed only for a 2x2 table					

There is enough statistical evidence from the sample that there is an association between gender and medical aid society chosen ($p<0.05$). In Table 3, results show that males are 85% more likely to prefer government schemes as compared to females (OR=1.85, 95% CI=1.44-2.38, p-value=0.001).

Table 4: Gender composition in each scheme type

		Scheme type			*Total*
		Low Premium scheme	*Medium Premium scheme*	*High premium scheme*	
Gender	Male	285	200	105	590 (58.1%)
	Female	205	135	85	425 (41.9%)
Total		490(48.3%)	335(33.0%)	190(18.7%)	1015

Table 5: Chi-Square Tests for gender compared to scheme type

	Value	*df*	*Asymp. Sig. (2-sided)*
Pearson Chi-Square	.982[a]	2	.612
Likelihood Ratio	.979	2	.613
Linear-by-Linear Association	.219	1	.640
N of Valid Cases	1015		
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 79.56.			

Results in Table 4 and Table 5 show that there is no association between gender and scheme type ($p>0.05$). This implies that in each scheme, claiming by clients is independent of whether they are males or females.

Table 6: Medical aid scheme compared to scheme type

		Schemetype			*Total*
		Low Premium scheme	*Medium Premium scheme*	*High premium scheme*	
MAS	Government scheme	215	175	130	520(51.2%)
	Private scheme	275	160	60	495(48.8%)
Total		490(48.3%)	335(33.0%)	190(18.7%)	1015

Table 7: Chi-Square Tests for medical aid scheme compared to scheme type

	Value	*df*	*Asymp. Sig. (2-sided)*
Pearson Chi-Square	33.212[a]	2	.000
Likelihood Ratio	33.829	2	.000
Linear-by-Linear Association	31.909	1	.000
N of Valid Cases	1015		
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 92.66.			

There is enough statistical evidence from the sample that there is an association between medical aid society and scheme type ($p<0.05$). Table 6 and Table 7 show that claiming was dependent on medical aid an individual had and scheme that they were enrolled in.

Cox regression model

Hazards for an individual given the explanatory variables are found using the Cox regression model. The aim of the model was to determine factors for a shorter time to a claim at a particular point in time. In the study, a Cox model was applied to the data using age, sex, medical aid societies (MAS), scheme type, and amount claimed as explanatory variables, and average time between claims as the dependent variable. The output is shown in Table 4.8. The number of observations is 1010 (n=1010).

Table 8: Estimates of variables in Cox regression model

Covariate	*b*	*exp(b)*	*se(b)*	*z*	*P*
Age	-0.0196508	0.9805410	0.0089510	-2.195	0.028 **
Male	0.1565140	1.1694272	0.0670201	2.335	0.020 **
Government	-0.0905288	0.9134480	0.0682929	-1.326	0.185
Medium premium	0.1488465	1.1604949	0.0723566	2.057	0.040 **
High premium	-0.4269528	0.6524943	0.0928022	-4.601	0.000 ***
Amount claimed	-0.0004320	0.9995681	0.0003574	-1.209	0.227

Key: ***-1%, **-5%, *-10% significant level.

In Table 8, regression coefficients are in second column (represented by **b**). There is need to understand that a positive coefficient implies increased hazard ratio while a negative coefficient implies a reduced hazard ratio. This means positive coefficient means time to a claim is reduced while when negative time to a claim is increased. Age has a negative coefficient and this implies risk of making a claim will decrease or average time to a claim will increase with increase in age. This implies that as individuals grow older, on average, they take a longer time to make a claim. This is due to the fact that, in Zimbabwe, senior citizens get free medication for less complicated ailments. The claims are just for critical ailments while younger patients do not have that privilege. Males, on average, have a shorter time to a claim as compared to females. Males are 16.9% ($p<0.05$) more likely to claim than females, holding other variables constant.

The variable MAS (medical aid society) is encoded as either private or government. The coefficient for MAS = -0.0905288 indicates that the average time to a claim for individuals from government schemes is longer as compared to their counterparts in private schemes. This is most likely due to the fact that government schemes are not adequate. If claims are made, out of pocket payments are highly likely to occur. The more claims you make, the more out of pocket payments you make. This discourages frequent claims by people having government schemes. Clients from government schemes make claims only during critical circumstances.

Third column in Table 8 specifies the estimate of hazard (**exp(b)**). For example, exp (-0.4269528) = 0.6525, meaning clients in high premium schemes are 65% less likely to make claims in particular time period as compared to clients in low premium schemes. This means clients in lower premium scheme claim more frequently in comparison to clients in high premium scheme. Clients in high premium schemes have a longer time to a claim while those in medium premium schemes a shorter time to a claim as compared to those in low premium schemes, holding other variables constant.

The Wald statistic values are given under the column marked "**z**" and these statistics correspond to the ratio of each regression coefficient and to its standard error (**z = b=se (b)**). To assess whether the coefficient of a variable in the model is statistically significantly different from zero, the Wald statistic is used. Using information in Table 8 above, it can be concluded that the variables age, sex and scheme type are statistically significant coefficients at 5% level of significance.

Hence from Table 8 the inferred Cox regression model for the study is;

$$h(t) = h_0(t)\exp(-0.01965x_1 + 0.1565x_2 + 0.1488x_3 - 0.42695x_4) \quad \text{Equation 4.1}$$

where:

$h(t)$ = expected hazard

$h_0(t)$ = baseline hazard function

x_1 = age

x_2 = gender

x_3 = scheme type 2

x_4 = scheme type 3

After inclusion of interactions, amount and scheme type are statistically significant. Sex, MAS and age become insignificant in influencing time to a claim. From Table 9, regardless of gender, scheme type has an influence on

Table 9: Cox regression model with interactions

	B	*SE*	*Wald*	*df*	*Sig.*	*Exp(B)*	*95.0% CI for Exp(B)*	
							Lower	*Upper*
Gender	.045	.162	.077	1	.782	1.046	.761	1.438
Age	-.032	.023	1.951	1	.162	.968	.925	1.013
MAS	-.162	.189	.735	1	.391	.850	.587	1.232
Scheme type			27.635	2	.000***			
Scheme type (LPS)	1.998	1.451	1.897	1	.168	7.374	.429	126.665
Scheme type (MPS)	-3.760	1.504	6.253	1	.012**	.023	.001	.444
Amount claimed	-.003	.001	21.197	1	.000***	.997	.996	.998
Age *Scheme type			25.233	2	.000***			
Age *Scheme type (LPS)	-.041	.027	2.341	1	.126	.960	.911	1.012
Age *Scheme type (MPS)	.062	.028	4.980	1	.026**	1.064	1.008	1.123
MAS*Scheme type			43.211	2	.000***			
MAS *Scheme type (LPS)	-.081	.215	.142	1	.707	.922	.606	1.405
MAS*Scheme type(MPS)	.916	.223	16.876	1	.000***	2.498	1.614	3.867
Amount* Scheme type			23.418	2	.000***			
Amount* Scheme type (LPS)	.003	.001	23.334	1	.000***	1.003	1.002	1.005
Amount *Scheme type (MPS)	.003	.001	15.621	1	.000***	1.003	1.001	1.004
Scheme type* Sex			.133	2	.936			
Scheme type (LPS)*Sex	.069	.191	.132	1	.716	1.072	.738	1.558
Scheme type (MPS)*Sex	.046	.206	.050	1	.824	1.047	.699	1.569

average time to a claim. Those with a medium premium scheme have shorter average time to a claim compared to other scheme types (HR=2.498, 95% CI= 1.614-3.867, p<0.05).

In spite of age, scheme type has an influence on average time to a claim. Clients with medium premium schemes have shorter average time to a claim irrespective of their age (HR=1.064, 95% CI=1.008-1.123, p<0.05). Amount claimed has become significant after interactions. The results show that as the amount claimed increases, average time to a claim increases (HR=0.997, 95% CI=0.996-0.998, p<0.05). This indicates that clients who claim large sums of amounts take more time to place in a claim, on average, compared to those who make small claims.

Irrespective of scheme type, amount is influential in determining average time to a claim. The hazard is almost similar in all scheme types (HR=1.003, 95% CI=1.001-1.004 for MPS and 95% CI =1.002-1.005 for LPS and p<0.05 in both cases)

Assessing proportional hazards assumption

Global goodness of fit test for the Cox regression model is the Schoenfeld residuals, used to test the assumption of proportional hazard (Su and Tsai, 2005). It was found that the global test show presence of robust proportionality. This is evidenced by the large p-values, which is an indication of strong proportionality that are supported by small global test statistics.

The results for the global test show that gender and scheme type are statistically significant since the p-values are greater than 0.05. Therefore, we can assume the proportional hazards. Figure 1 below show five smoothed scaled Schoenfeld residual plots for significant and nonsignificant variables under analysis. The plots show that gender, amount claimed and scheme type are significant variables.

Checking for Outliers

Poorly predicted observations were of concern to check. To assess the magnitude of outliers, the deviance residual plots were used. This methodology detects how well the data describes the data when there is inclusion of explanatory variables in the final model. If there is no pattern form then the model best describes the data but if the points are not symmetrically distributed around zero, then the model does not explain the data satisfactorily. It is worth noting

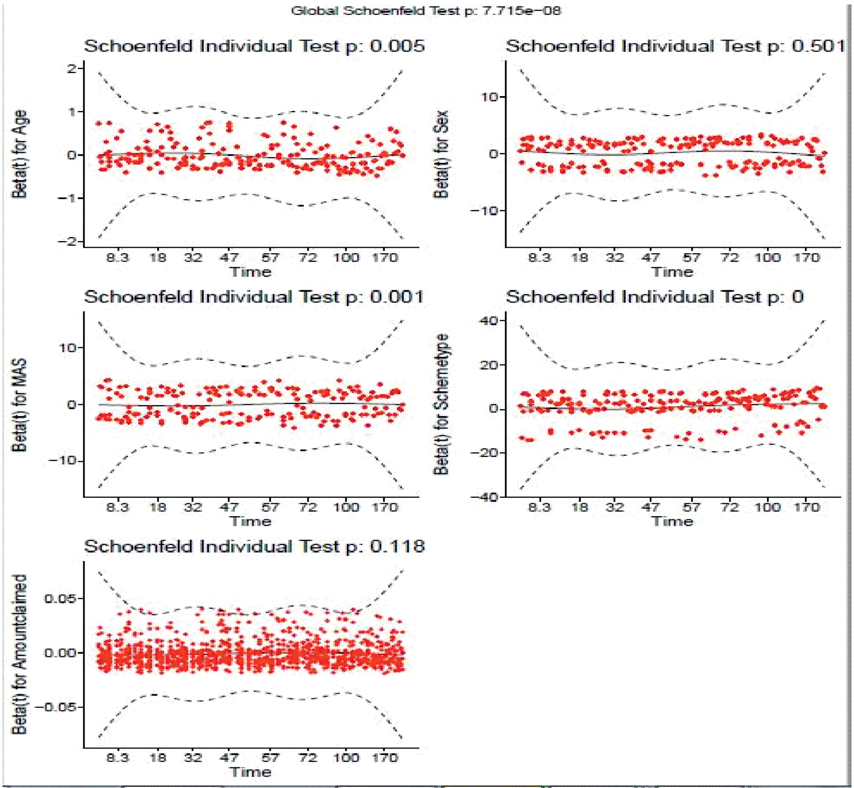

Figure 1: Residual plots for explanatory variables against time.

that the deviance residual method has, as according to Dessai and Patil (2019), a weakness of being a subjective method.

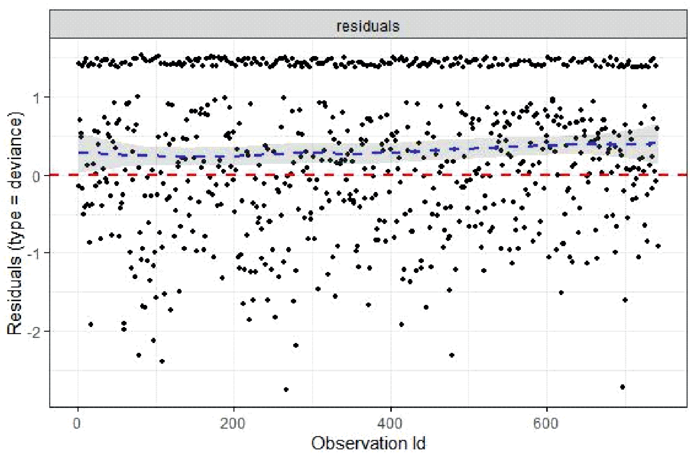

Figure 2: Detecting outliers using the deviance residuals plot

In Figure 2 and 3, the blue line represents the smoothing line of the plot with the dark shade illustrating the confidence interval.

Checking non- linearity

The functional form of the covariates was being investigated here. To assess the functional form of the explanatory variables, martingale residuals were plotted. If there is a desired linear or close to linear plotting of the functional form, then the assumption is not violated otherwise, if there is no linearity, the assumption will have been violated. If there is evidence of violation of the assumption, it is highly recommended that data transformation techniques (eg. log, square root, square and other transformations) be conducted. Figure 3 show that the martingale residual plots show a functional form for the explanatory variables such as age, gender, medical aid society and scheme type that seem close to linear.

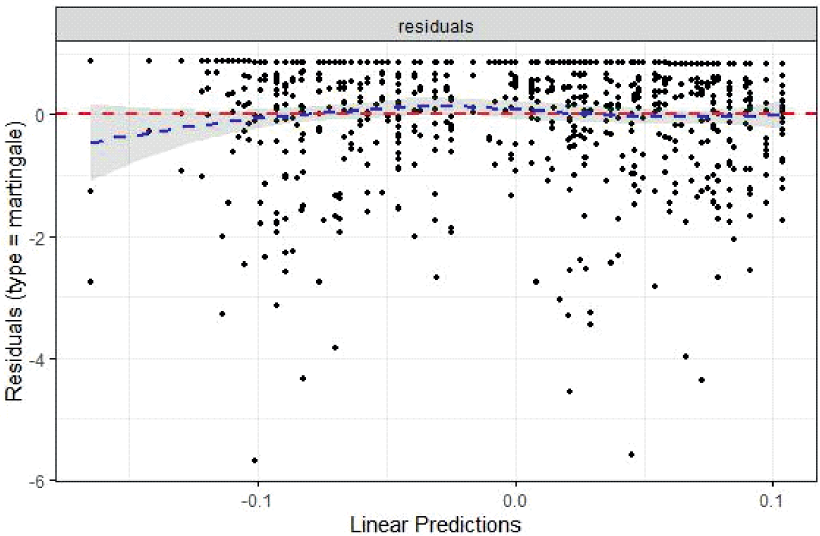

Figure 3: Martingale residual plot

Discussion of results

Several statistical models for claims have been proposed in recent years, and now the list of available approaches is much wider. Nevertheless, there is no unique model that is able to deal with all the problems that can arise in the analysis of claims data. Literature has shown that actuarial, econometric, and both non-parametric and parametric survival models are commonly used in many claims researches. Thus, the final decision is dependent on the type and layout of the study.

In this study, the researcher proposed the use of Cox regression model to examine the effect of explanatory variables on time to claim. This model allows for formal testing of the proportional hazard assumption, outliers and none linearity using residuals. Table 8 shows that there are four significant variables that are to be included in the model. These have their p-values less than 0.05. In addition, the Cox model was conducted to establish the influential significant factors. The hazard ratio has been evaluated for each significant factor; age, sex and scheme type. It was established that these were the variables influential in determining length of time to a claim and are essential in estimating the hazard rate.

Like in other researches by Mhere (2013), Nketiah-Amponsah and Arthur (2013), Sari *et al* (2019), Freeman (2011) and others on determinants of factors for participation in health insurance, age (p-value=0.0281) is a significant factor in determining time to a claim. This likely due to the fact that as one grows older, one becomes more prone to medical conditions. On the contrary to what Kong and Kim (2020) found that more medical aid beneficiaries were aged 60 or over, in this research the coefficient for age (-0.0196508) is negative indicating that as individuals reach 65 or more, they are likely to claim less. This is due to the fact that in Zimbabwe, the elderly population, normally called senior citizens go to government clinics or hospitals for free healthcare services and medical aid payments are supported by employers and pensioners cannot afford the payment without the employer's support. Irrefutably, holding all other explanatory variables constant, age proved to have a significant impact on time to a claim.

Gender was found to be influential in determining time to a claim. Unlike what Kong and Kim (2020) found out in their research on medical aid beneficiaries in Korea that women tended to have higher rates of claim, the current study found out that men have a shorter time to a claim as compared to women, males are 16.9% more likely to make a claim as compared to females. This means males claim more frequently or benefit more frequently from medical aid policies as compared to their female counterparts. There is low participation of women in health insurance in Zimbabwe. As according to Sipsma *et al* (2013) lack of empowerment for women is the likely reason for underutilisation of healthcare services. Mhere (2013) indicates that 91% of women in Zimbabwe do not have health insurance. This could be the likely reason why in Zimbabwe males tend to have a shorter time to a claim as compared to women.

Medium premium subscribers tend to have a shorter time to a claim as compared to low premium subscribers. This is in line with results found by Jiang *et al* (2018), Xianzhj (2018), Nosratnejad and Shami (2017), Bertranou (1998) and Muchabaiwa (2012) found. Medium premium subscribers could pay less out of pocket and hence influence the rate at which they claim while low premium schemes discourage shorter times to claims as they are not fully adequate and encourage more out of pocket payment which eventually discourage frequent claims unless really critical. Shorter time to a claim means high utilisation of the healthcare services.

On another note, high premium subscribers had longer times to a claim compared to low premium subscribers. High premium subscribers were 65% less likely to make a claim as compared to low premium subscribers, holding other variables constant. This shows that clients in low premium schemes took a shorter time to a claim as compared to clients in high premium schemes. It might be assumed that high premium subscribers are expected to use healthcare services more frequently compared to low premium subscribers. This is in line with previous researches by Zhang *et al* (2020), Geil *et al* (1997), Riphahn *et al* (2003) and Jurges (2009). Our result could be due to the fact that high premium subscribers are more likely using private hospitals, as found by Esmailnasab (2014), compared to public hospitals where data was collected from.

Medical aid society (MAS) and amount claimed were found to be not significant in influencing time to a claim. This implies that, irrespective of other factors, amount claimed does not influence time to a claim. As found in previous researches by Wammes *et al* (2017) and Kalseth and Halvorsen (2020), health costs in terms of age roughly follow the same pattern. As one grows older, the costs associated to his health are likely to increase. Cost on use of health care services on the other hand tend to be stable or decrease with increase in age. This necessitates the fact that most interventions need not to exclusively look at elderly people but all eligible members.

Medical aid society was insignificant in influencing time to a claim. As found by Garg *et al* (2019) enrolment under Publicly Funded Health Insurance (PFHI) was not associated with increase in healthcare utilisation. In this research, whether you are in a private medical aid society or government or public medical society does not reduce time to a claim which implies frequent use of healthcare services.

CONCLUSIONS AND RECOMMENDATIONS

Medical aid societies (MAS) and amount claimed were not significant factors that affect time to claim. This implies that they are not influential in shortening time to a claim. The covariates Age, Sex and Scheme types are highly significant and influential in determining time to making claim. After introduction of interactions, amount claimed and scheme types were found to be significant in shortening time to a claim.

Scheme type and sex interaction was not significant. This implied that, irrespective of one's gender, scheme type was highly influential in shortening time to a claim. Scheme type and amount interaction, scheme type and age interaction and finally scheme type and MAS interaction were significant in shortening time to a claim.

In each of the significant interactions, it was noted that clients in medium premium schemes had a shorter time to a claim compared to clients in other scheme types. The study therefore recommends that, the insurance companies should know the expected time to a claim for their clients in order to optimize benefits packages to cover clients' medical expenses, yet remain within sustainable operating limits. The research also proposes that medical aid companies have different types of scheme for different types of clients.

References

Association of Healthcare Funders of Zimbabwe. (2008). 'Funding and providing healthcare in Zimbabwe,' *BHF Annual Conference, 13-16 July 2008, Durban. Board of Healthcare Funders: Johannesburg.*

Banke-Thomas, O.E., Banke-Thomas, A.O. & Ameh, C.A.(2017). Factors influencing utilisation of maternal health services by adolescent mothers in Low-and middle-income countries: a systematic review. *BMC Pregnancy Childbirth* 17, 65 (2017). https://doi.org/10.1186/s12884-017-1246-3

Bertranou, F. (2010). Health Care Services Utilization and Health Insurance Coverage: Evidence from Argentina. *Revista de Análisis Económico – Economic Analysis Review, 13*(2), 25-52. Retrieved from https://www.rae-ear.org/index.php/rae/article/view/130/248.

Cekici, E. M., Aydogan, H., Eti, S. (2018). Forecasting of claim by combining probability distributions. Journal of Business Economics and Finance (JBEF), V.7(3), p.319-324. http://doi.org/10.17261/Pressacademia.2018.959.

CIMAS (2009) 'Private hospital benefit package,' *Information Bulletin. CIMAS: Harare.*

COX, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34 (2), 187–202.

Dalal, A., Zimmerman, E., Magnoni, B and Matul, M. (2014). Is there value in microinsurance?, *Client Value Brief No. 1 (Geneva, ILO)*

Dessai, S and Patil, V.(2019). Testing and interpreting assumptions of COX regression analysis. *Cancer Research Statistics and Treatment,* 2019; 2:108-11.

Esmailnasab, N., Hassanzadeh, J., Rezaeian, S and Barkhordari, M. (2014). Use of Health Care Services and Associated Factors among Women. *Iranian journal of public health*, *43*(1), 70–78.

Garg, S., Chowdhury, S and Sundararaman, T. (2019).Utilisation and financial protection for hospital care under publicly funded health insurance in three states in Southern India. *BMC Health Services Research* 19, 1004 (2019). https://doi.org/10.1186/s12913-019-4849-8

Gascon, G. M., Notz, W. I., and Waller, B.A.(2016). Fitting probability distributions to health plan claims data: Implications for enhanced case management. Journal of propagations in probability and statistics. 16(1), 1-12.

Geil P, Million A, Rotte R and Zimmermann K.F.(1997). Economic incentives and hospitalization in Germany. *Journal of Applied Economics.* 1997;12:295–311.

Gustafsson, E. (2009). Customer duration in non-life insurance industry. Mathematical statistics, dept. of mathematics, Stockholm University, Sweden.

Jennings, D.E. (1986). Outliers and Residual Distributions in Logistic Regression, *Journal of the American Statistical Association*, 81(396), 987-990, doi: 10.1080/01621459.1986.10478362

Jiang, M., Yang, G., Fang, L., Wan, J., Yang, Y and Wang, Y. (2018). Factors associated with healthcare utilization among community-dwelling elderly in Shanghai, China. *PLoS ONE 13(12): e0207646*. https://doi.org/10.1371/journal.pone.0207646

Jürges, H.(2009). Health insurance status and physician behavior in Germany. Schmollers Jahr. 2009; 129: 297.

Kalseth, J., and Halvorsen, T. (2020). Health and care service utilisation and cost over the life-span: a descriptive analysis of population data. *BMC Health Services Research* 20, 435 (2020). https://doi.org/10.1186/s12913-020-05295-2

Kong, N.Y., and Kim, D.H. (2020). Factors influencing health care use by health insurance subscribers and medical aid beneficiaries: a study based on data from the Korea welfare panel study database. *BMC Public Health* 20, 1133 (2020). https://doi.org/10.1186/s12889-020-09073-x

Konrad, R., Zhang, W., Bjarndóttir, M., & Proaño, R. (2019). Key considerations when using health insurance claims data in advanced data analyses: an experience report. *Health systems (Basingstoke, England)*, *9*(4), 317–325. https://doi.org/10.1080/20476965.2019.1581433

Mashasha, M., Mutize, P., and Mazunga, F.(2022). Distibution and Pattern of an insurance health claim system: A time series approach. *Tanzania Journal of Science* 48 (1), 13-21, DOI: https://dx.doi.org/10.4314/tjs.v48i1.2

Mhere, F. (2013). Health insurance determinants in Zimbabwe: Case of Gweru Urban. *Journal of Applied Business and Economics,* vol. 14(2), 62-79.

Muchabaiwa, Lazarus and Mazambani, D. and Chigusiwa, Lloyd and Bindu, Samuel and Mudavanhu, V. (2012).Determinants of Maternal Healthcare Utilization in Zimbabwe. *International Journal of Economic Sciences and Applied Research,* Volume 5, Issue 2, pp. 145-162, Available at SSRN: https://ssrn.com/abstract=2155627

Nketiah-Amponsah, E., and Arthur, E. (2013). "Choice of delivery facility among expectant mothersin Ghana does access to health insurance matter?"*Journal of Health Management*,15(4), 509–524.

Nosratnejad, S and Shami, E. (2017). Health insurance and the utilization of health care: a systematic review. *BMJ Open,*7(1). doi: 10.1136/bmjopen-2016-015415.70

Qian, D.,Lucas, H., Chen, J., Xu, L and Zhang, Y.(2010).Determinants of the use of different types of health care provider in urban China: A tracer illness study of URTI. *Health Policy*. doi:10.1016/j.healthpol.2010.06.014

Rendek, K., Holtz, J and Fonseca, C. (2014). The moment of truth: Claims management in microinsurance. *Microinsurance Paper No. 28 (Geneva, ILO).*

Riphahn, R.T., Wambach, A and Million, A.(2003). Incentive effects in the demand for health care: a bivariate panel count data estimation. *Journal of Applied Economics*. 2003;18:387–405.

Sello, D.A., Serfontein, J.H.P., Lubbe, M.S and Dambisya, Y.M. (2012). Factors influencing access to pharmaceutical services in underserviced areas of the West Rand District, Gauteng Province, South Africa*, Health SA Gesondheid* 17(1), Art. #609, 8 pages. http://dx.doi.org/10.4102/hsag.v17i1.609

Sipsma, H., Callands, T. A., Bradley, E., Harris, B., Johnson, B and Hansen, N. B. (2013). Healthcare utilisation and empowerment among women in Liberia. *Journal of epidemiology and community health*, *67*(11), 953–959. https://doi.org/10.1136/jech-2013-202647

Su, X and Tsai, C.L. (2005). Tree-augmented cox proportional hazards models. *Biostatistics,* 6 (3), 486–499.

Wammes, J.J.G., Tanke, M., Jonkers, W.,Westert, G.P., Van der Wees, P and Jeurissen, P.P.T .(2017). Characteristics and healthcare utilisation patterns of high-cost beneficiaries in the Netherlands: a cross-sectional claims database study. *BMJ Open.* 7(11). doi: 10.1136/bmjopen-2017-017775

Xianzhi, F., Nan, S., Fei, X., Jin, L., Qixin, T., Junjian, H., Dongdong, W and Changqing, S. (2018) Influencing factors of inequity in health services utilization among the elderly in China. *Int J Equity Health* 17, 144 (2018). https://doi.org/10.1186/s12939-018-0861-6

Xie, Y., Schreier, G., Hoy, M., Liu, Y., Neubauer, S., Chang, D. C. W., Redmond, S. J., Lovell, N. H. (2016). Analyzing health insurance claims on different timescales to predict days in hospital, *Journal of Biomedical Informatics*, 60(2016), 187-196, https://doi.org/10.1016/j.jbi.2016.01.002.

Xue, X., Xie, X., Gunter, M., Rohan, T. E., Wassertheil-Smoller, S., Ho, G. Y., Cirillo, D., Yu, H and Strickler, H. D. (2013). Testing the proportional hazards assumption in case-cohort analysis. *BMC medical research methodology*, 13 (1), 88.

Zhang, C., Fu, C., Song, Y., Feng, R., Wu, X. and Li, Y. (2020). Utilization of public health care by people with private health insurance: a systematic review and meta-analysis. *BMC Public Health* 20, 1153 (2020). https://doi.org/10.1186/s12889-020-08861-9

Zhang, N. J. (2010). Regression survival analysis with dependent censoring and a change point for the hazard rate: With application to the impact of the Gramm-Leach-Bliley Act to insurance companies' survival. *Ph.D. thesis.*