ISSN 2278-3091
Volume 9, No.1.4, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse1491.42020.pdf
https://doi.org/10.30534/ijatcse/2020/1491.42020

Evaluation of Risk Factor for Cancer Insurance using R Analytics

Nurin Haniah Asmuni, Norkhairunnisa Mohamed Redzwan, Norazliani MdLazam, Nordiana Rosdi
UniversitiTeknologi MARA, Malaysia, nurin@tmsk.uitm.edu.my
UniversitiTeknologi MARA, Malaysia, khairunnisa@tmsk.uitm.edu.my
UniversitiTeknologi MARA, Malaysia, norazliani@tmsk.uitm.edu.my
UniversitiTeknologi MARA, Malaysia, nordiana_rosdi@tmsk.uitm.edu.my

ABSTRACT

Some faulty genes causing cancer disease are inherited and some are caused by lifestyle factors. This research focuses on the implication of utilizing a family history information as a risk factor in estimating cancer incidence rate and the cost of insurance. The Maximum Likelihood Estimator (MLE) approach is applied to estimate the cancer incidence rates, while the transition probabilities correspond to each event under the insurance coverage are estimated using a continuous Markov Model. This study emphasizes on the application of R Analytics in assessing the risk factor and the graduation procedure of the estimated incidence rates. Our results show that the insurance cost for people with family history of cancer is significantly higher in comparison to those without family history of the disease especially for females. Overall, the cost of cancer insurance for both gender categories increases as age increases.

Key words : Cancer, family history, Markov, risk factor, R Analytics.

1. INTRODUCTION

Cancer is one of the leading critical illnesses in many countries around the world. Cancer is a disease caused by the uncontrolled growth and spread of malignant cells within the body called tumor. According to the World Health Organization, there were more than a quarter of deaths attributable to cancer in many countries in year 2000 and it is predicted that in 2020, cancer rate will increase by 50% to 15 million of new cancer cases [1].

Modifiable lifestyle factors such as smoking, alcohol use and diet quality may affect the incidence of critical illness. The study by [2] concludes that healthy lifestyle is associated with lower risk of cancer. Apart from lifestyle factors, previous studies have shown the importance of genetics effects as a risk factor in determining the rate of critical illness [3], [4]. The term “genetics” is defined as the study of heredity and the variation of inherited characteristics.

In insurance sector, the only genetic information acquired during underwriting stage in some countries is family history as insurers’ access to genetic test results is usually confidential [5]. The possibility to use genetic information in insurance industry has been controversial ever since genetic tests development in the 1990s. According to [6], most life and health insurance proposal forms require applicants’ family history details and, as with existing conditions, the failure to disclose a known family history of genetic disorder would be viewed as fraud and may discredit the policy. The family history information required is typically about parents and siblings in relation to the occurrence of cancer, heart disease, hypertension, stroke and kidney disease. [7] suggests that it is common for an insurer to ask for the age at diagnose of disease for any affected relatives of the insurance applicant.

In this paper, we intend to study how significant the implication of using a family history information as a risk factor on the cost of cancer insurance. The aim of our study is to determine whether the existence of family history impacts significantly on the insurance cost. Though risk classification topic in insurance industry has been regularly discussed in past actuarial literature [8], [9], [10], consideration of risk factors for insurance premium has yet to be explored in Malaysia which is important to promote premium transparency for consumers.

2. LITERATURE REVIEW

Cancer is the second leading cause of death globally and 8.8 million of deaths are attributable to cancer in 2015. According to [11], the survival rate of cancer patients in Malaysia depends on the cancer type where the lowest survival rate is 11% for patients of lung cancer with diagnosis year of 2007 to 2011. The Malaysia’s Minister of Health, in his recent statement underscores the need to optimally plan for cancer related treatment especially in private hospitals to save on treatment cost [12].

Cancer imposes a tremendous financial burden to cancer patients and their family members. Based on a study by [13], the total cost of colorectal cancer management of new cases in 2012 is RM107,699,768. An insurance coverage for cancer is an ideal product to reduce financial burden associated with cancer treatments. For instance, some insurers offer a lump sum benefit payment specifically upon the diagnosis of cancer

disease. While other products like mySalam which was introduced by the government of Malaysia in 2019, provides a lump sum benefit upon the diagnosis of 47 types of critical illness [14]. For such coverage to be offered in the market, an accurate estimation of cancer incidence rates is important for insurance provider to avoid the underestimation of such risk.

Recently, there have been an increasing use of data analytic techniques in the healthcare studies to estimate a particular disease rate [15], [16], [17]. In this study, we observe the cancer incidence rates pattern among hospital patients according to their age, gender and family history record. Here, the analysis is performed using hybrid model consisting of data analytics and traditional actuarial techniques.

In the process of determining an appropriate insurance pricing for non-life insurance, risk classification is crucial to avoid over subsidizing the group of insured with high risk. [18] states that private insurance is based on the principle of mutuality in which it is voluntary, applicants are grouped by their respective risk, and premium differs between these groups reflecting the likelihood of claims. [8] indicates that insurers normally use risk factors which considerably affect specific risks that are associated to the coverage they are offering, allowing them to classify a group of policyholder according to the risk levels. This information is useful for setting diverse rates relevant to members in each risk class.

According to [19], the population can be divided into persons with no family history who are not at risk, and persons with a family history of whom have a higher chance to carry the risk. While the use of age and gender as risk factors is common in the insurance industry, little is known on how the family history information could affect the insurance pricing. Thus, this study aims to explore this area further within Malaysian population.

3. METHODOLOGY

Our analysis is conducted by collecting secondary data from Hospital Kuala Lumpur (HKL) patients' record. This data includes the information of cancer patients and medical checkup patients which were registered in year 2007. The main variables collected consists of the date of birth, date of cancer diagnosis, gender, age, date of death (if death occurs in 2007) and family history of cancer diagnosis. The range of age considered in the data is between 20 to 75 years old for both male and female groups. A total number of 1033 observations has been collected based on the patients' profile.

3.1 Cox Proportional Hazard Regression in R

First of all, we evaluate the significance of several risk factors based on the cancer patient's profile data namely age, gender and family history information. The objective is to determine if the likelihood of cancer incidence in the sample can be explained by the risk profile of the patient. Cox Proportional

```
x <- read.csv(file="data_cancer.csv", header = TRUE,
      sep = ",",dec = "." )
head(x)

#List of variables
# exposure: total waiting time in healthy state in days
# status: 1=censored, 2=diagnosed with cancer
# age: age in years
# gender: 1=male, 2=female
# family history (fh): 1=yes, 2=no

install.packages(c("survival", "survminer"))
install.packages("ggplot2")
install.packages("ggpubr")
library("survival")
library("survminer")
res.cox <- coxph(Surv(exposure, status) ~ age + gender + fh, data = x)
summary(res.cox)
```

Figure 1: Coding for Regression Analysis

Hazard Regression invented by [20] suits this purpose and this model has been widely applied in the population and medical research area. The regression model has the following form:

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) \quad (1)$$

where $h(t)$ is the expected hazard rate at time t and $h_0(t)$ is known as the baseline hazard which represents the hazard rate when all covariates X_1, X_2, X_3 are equal to zero. In this study, the covariates are $X_1 = age$, $X_2 = gender$ and $X_3 = familyhistory$. We compute the coefficients in the model by utilizing package survival and survminer in R Analytics for Statistical Computing Software as described in [21]. The coding for our regression analysis is presented in Figure 1.

3.2 Estimation of Cancer Incidence Rates

Based on the obtained data, the number of patients who are diagnosed with cancer from the total exposure will be observed. They are classified into two categories, with and without family history of cancer diagnosis. We apply the Maximum Likelihood Estimation (MLE) to estimate the cancer rate. [22] describes MLE as a standard method for parameter estimation and inference in statistics. Under Markov model's assumption, the lifetime function follows the exponential distribution. Let us consider a sample with n independent lives, the total likelihood function over all n lives is as follows:

$$L \propto \prod_{i=1}^{n} e^{-(t_i - r_i)\nu} \nu^C \quad (2)$$

Where L is the likelihood function from time r to t, C is the number of cancer patients with (or without) family history factor between time r and t, and ν is the instantaneous incidence rate of cancer. The equation is solved for ν by taking the log of the likelihood function.

$\ell = \ln L$

$\ell = C \ln \nu - \nu \sum_{i=1}^{n} (t_i - r_i)$

By equating the derivative of the function to 0, we obtain:

$$\hat{v}_x = \frac{C_x}{\sum_{i=1}^{n}(t_i - r_i)} \quad (3)$$

Equation (3) is the ratio of C_x to the exact exposure of the sample, which is the sample central rate where sample is in group of age x. Thus this will be used to estimate $\hat{v}_x^{FH}$ which represents the incidence rate of cancer for people with family history of the disease and $\hat{v}_x^{NFH}$ which represents the incidence rate of cancer for people without family history of the disease.

3.2.1 The Graduation Procedure

Once the crude cancer incidence rates are estimated, we apply two graduation techniques and choose the best fitting method for our cost calculation in the next subsection. This step is useful for converting the group incidence rate to single age incidence rate. The first graduation technique has been used extensively in the survival analysis study which is known as the Gompertz model[23], [24]. The Gompertz law states that $\log(v_x)$ is linear in age as follows:

$$\log(v_x) = a + bx \quad (4)$$

where parameter a and b are solved by minimizing the sum of squared errors between the crude incidence rate and the graduated incidence rate. We solve the parameters using Solver application in Excel.

The second graduation technique is called the cubic spline method which is a piecewise polynomial function consists of pieces of third-order polynomial, smoothly joint at some points called knots. According to [25], splines are regularly used for building explanatory models in clinical research. Though solving the function is more mathematically challenging, there is a ready package in R Analytics which can be utilized to solve the function numerically. Here, the following R coding in Figure 2 is applied to graduate the crude incidence rates using cubic spline method.

```
x <- scan(file="cancerrate_male.csv", what = "character", skip = 1,
      sep = ',')
x.mat <- matrix(x, ncol = 3, byrow = TRUE)
cancerx <- x.mat[2:12, 2:3]
cancerx <- matrix(as.numeric(cancerx), nrow = 11, ncol = 2)
Age <- c(21, 26, 31,36,41,46,51,56,61,66,71)
Status <- c("FH","NFH")
dimnames(cancerx) = list(Age, Status)
cancerx[1,]
#spline fitting to find single age cancer incidence rate
i <- 1:2
single <- sapply(i, function(x) spline(Age[1:11],cancerx[1:11,x], n = 51))
single[1:4]
```

Figure 2: R Coding Using Cubic Spline Method

3.3 The Cost of Cancer Insurance

The calculation of insurance premium which consists of benefit payment in more than one health status event requires a model that can incorporate several health states rather than common survival model. Past actuarial literature have extensively applied the multiple states Markov model for this purpose [8], [19], [26]. In this paper, we apply a continuous 3-state Markov model as illustrated in Figure 3.

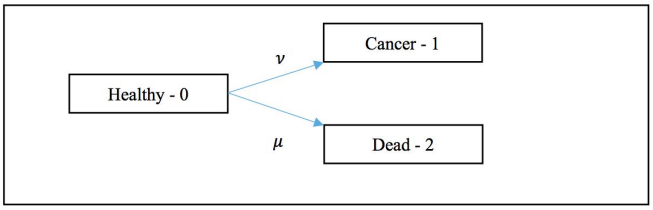

Figure 3: The 3-State Markov Model

This model assumed that the history of a life starts as a healthy policyholder at age x, with the possible transition shown by the arrows. State 1 in Figure 3 above represents the state of health of patients diagnosed with cancer at any stages. The main features of the model are as follows;

(a) No recovery is allowed from the cancer state. Since the payment of cancer benefit is payable only once in the event of cancer diagnosis, the transition rate from cancer state to healthy state is not applicable in our model.

(b) Mortality rates for healthy individuals is assumed to follow the Malaysian population mortality table obtained from the Department of Statistics, Malaysia. We use the 2016 mortality rates to match the cancer incidence rates that are calibrated from the National Cancer Registry Report 2012-2016 [27]. The force of mortality in this model is denoted as μ.

(c) The model follows an assumption by [28] where all transition probabilities are assumed to be stationary over time.

There are three transition probabilities which correspond to the events; (i) Healthy to Cancer, (ii) Healthy to Dead, (iii) Healthy to Healthy. The notation P_x^{ij} will be used in the following derivation steps, which is defined as the transition probability of people currently aged x who make a transition from state i to j in one year. Our estimation is explained further below:

Healthy to Cancer, $\boldsymbol{P_x^{01}}$: Suppose an insured makes a transition to the Cancer state (1) at time $x + t + dt$. We list all states where transition is possible to happen at time $x + t$, i.e. just before time $x + t + dt$. For P_x^{01}, the final Cancer state (1) can be reached from the Healthy state (0) and the Cancer state (1) as illustrated in Figure 4.

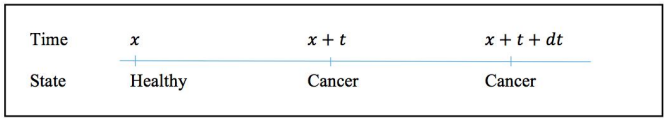

Figure 4: The Timeline of Events correspond to Transition Probabilities

The Kolmogorov Forward Equation for P_x^{01} is derived as follows:

$$ {}_{t+dt}P_x^{01} = {}_tP_x^{01} \times 1 + {}_tP_x^{00} \times (\nu dt + o(dt)) $$

The notation o can be defined as a random variable for a small error in the function which is closed to 0. Rearranging we obtain the following:

$$ \frac{{}_{t+dt}P_x^{01} - {}_tP_x^{01}}{dt} = \nu \, {}_tP_x^{00} + \frac{o(dt)}{dt} $$

Hence, taking the limit $dt \rightarrow 0$ we get,

$$ \frac{\partial}{\partial t} \, {}_tP_x^{01} = \nu \, . \, {}_tP_x^{00} \qquad (5) $$

Healthy to Dead, ${}_tP_x^{02}$: We can apply the same argument to ${}_tP_x^{02}$. We use the symmetry of Figure 3 written in the form of Kolmogorov forward equation for ${}_tP_x^{02}$ as follows:

$$ \frac{\partial}{\partial t} \, {}_tP_x^{02} = \mu \, . \, {}_tP_x^{00} \qquad (6) $$

Healthy to Healthy, ${}_tP_x^{00}$: Finally, for the probability of staying in healthy state (0), it is impossible to return the state 0 if one leaves the state, so ${}_tP_x^{00} = {}_tP_x^{\overline{00}}$ where the bar indicates that the person stays at state 0 for the whole time x to $x + t + dt$. The probability of making transition out of the healthy state in time dt is

μdt + vdt + o (dt)

Hence, the probability that the person remains in the healthy state for time dt is

1 - μdt - vdt + o (dt)

Thus, we obtain the following:

$$ {}_{t+dt}P_x^{\overline{00}} = {}_tP_x^{\overline{00}} \times (1 - \mu dt - \mathrm{v}dt + \mathrm{o}\,(dt) $$

Rearranging and letting $dt \rightarrow 0$ we get,

$$ \frac{\partial}{\partial t} \, {}_tP_x^{\overline{00}} = -(\mu + \mathrm{v}) \, . \, {}_tP_x^{\overline{00}} \qquad (7) $$

Putting ${}_tP_x^{00} = {}_tP_x^{\overline{00}}$, this is the Kolmogorov equation for ${}_tP_x^{00}$. In the present case, to solve this Kolmogorov equation, we rearrange as follows:

$$ \frac{\frac{\partial}{\partial t} \, {}_tP_x^{00}}{{}_tP_x^{00}} = -(\mu + \mathrm{v}) \longrightarrow \frac{\partial}{\partial t} log \, {}_tP_x^{00} = -(\mu + \mathrm{v}) $$

Then, integrating the above function with respect to t, we can solve for ${}_tP_x^{00}$:

$$ {}_tP_x^{00} = \exp\,[-(\mu + \mathrm{v})\, t] \qquad (8) $$

Now that we know ${}_tP_x^{00}$, we can solve for ${}_tP_x^{01}$ and ${}_tP_x^{02}$. Substitute Equation (8) to (5), we get,

$$ \frac{\partial}{\partial t} \, {}_tP_x^{01} = \nu \exp\,[-(\mu + \nu)\, t] $$

Integrating the above function from 0 to t gives us the following:

$$ {}_tP_x^{01} = \frac{\nu}{\mu+\nu} (1 - \exp\,[-(\mu + \nu)\, t]) \qquad (9) $$

And by symmetry we also get,

$$ {}_tP_x^{02} = \frac{\mu}{\mu+\nu} (1 - \exp\,[-(\mu + \nu)\, t]) \qquad (10) $$

3.3.1 Net Premium Calculation

For illustration purpose, we adopt the structure of cancer insurance product available in the Malaysia's insurance market. The sum assured is assumed to be RM 50,000 for 1 unit of contract. In calculating the actuarial present value of the benefits, the assumption of risk free interest rate of 4% per annum is applied. We compare this basic cost of insurance if the insured had undisclosed the family history information, as opposed to the cost calculated with family history information included as a risk factor. The insurance plan benefit is described below:

- Sum assured (RM): 50,000
- Benefit upon diagnosis of cancer: 100% of sum assured
- Compassionate benefit upon death: 10% of sum assured
- Coverage term: up to 70 years old.

The actuarial present value of net single premium for the benefit plan is presented in Equation (11) as follows:

$$ A_x^{(k)} = SA \times \{\sum_{t=0}^{70-x} \left(\frac{1}{1+i}\right)^{t+1} \left({}_tP_x^{00}[100\% \, {}^kP_{x+t}^{01} + 10\%. \, {}^kP_{x+t}^{02}]\right)\} \qquad (11) $$

where SA is the sum assured and i is the interest rate factor. ${}_t^kP_x^{ij}$ is the transition probability for a person aged x with family history of cancer $[k = FH]$ or without family history of cancer $[k = NFH]$, who makes a transition from state i to j in time t. If t is not stated in the notation, the transition probability is for one year.

4. FINDINGS

4.1 Risk Factor Evaluation

In this subsection, we present the Cox Proportional Hazard regression results where three covariates are applied namely age, gender and family history information. Our result is shown in Table 1.

Table 1: Cox Proportional Hazard Regression Analysis Result

Covariate	*Coefficients*	*Exp (Coef)*	*SE (Coef)*	*z-stat*	*P-value*
Age	0.005346	1.005360	0.002979	1.795	0.0727
Gender	0.103507	1.109054	0.087303	1.186	0.2358
Family History	-0.206650	0.813305	0.088548	-2.334	0.0196

The regression analysis result shows that the most significant factor affecting the cancer hazard rate in our sample is family history information at significance level of 5% (refer to *P*-value column in Table 1). The regression coefficient gives an indicator of the hazard ratio for the second group relative to the first group. For instance, if we look at the coefficient for family history variable, the coefficient of -0.21 indicates that the hazard (risk of cancer incidence) is lower for people without family history of cancer relative to people with family history of cancer. While the exponent coefficient (*Exp(coef)*) value gives the effect magnitude of covariate which means the risk for people without family history of cancer is reduced by a factor of 0.81 or 19%.

In a similar way, we can interpret the exponent coefficient for other variables as well. Our result shows that for age variable, the risk of cancer incidence is higher as age increases, whereas for gender variable, the risk is higher for females relative to males by 10.9%.Following this regression result, we have a sound statistical basis to calculate the incidence rates and insurance cost by category of policyholder with and without family history of cancer.

Generally, the calculation of incidence rates can be done using the result of Cox Proportional Hazard Regression in this section. However, this will only provide the rate for this particular hospital sample. Thus, instead of using the result directly from the hospital sample, we calculate the proportion of patients with and without family history of cancer and use these data to calibrate to the incidence rate for the whole population obtained from the National Cancer Registry Report.

4.2 Cancer Incidence Rates

The first step in estimating the cancer incidence rate is to group the collected patients data by age, gender and group of k either the patient has a family history of cancer or not. Next, we calculate the total exposure as a fraction of year. It involves the information on the scheduled date of birth and date of diagnosed of each patient. The total exposure for cancer patient at each observed age, from 20 to 70, is the exact age $x + t$ at which the patient is diagnosed with cancer, minus with the age at which observation begins, $x + r$. The data is assumed to be left censored, where we assume that the first diagnosis of cancer for all patients follows the date of diagnosis in the hospital record. Hence, all patients are assumed to be healthy prior to the date of diagnosis.

The total exposure shall include the total number of population at risk rather than just the cancer patients. Thus, we calibrate our data to the incidence rates obtained from the National Cancer Registry Report 2012-2016 [27]. In particular, the total exposure per person with cancer calculated from the hospital record is multiplied with the number of cancer cases calculated from the National Cancer Registry Report. Similarly, the proportion of population with family history of cancer follows the hospital record, and the proportion is multiplied with the total number of population at risk according to the National Cancer Registry Report. Then, the incidence rates of cancer for both categories with and without family history of cancer are estimated using Equation (3). The fitted incidence rates for male category with and without family history of cancer are shown in Figure 5(a) and 5(b) respectively. Similarly, the fitted incidence rates for female category are shown in Figure 6(a) and 6(b).

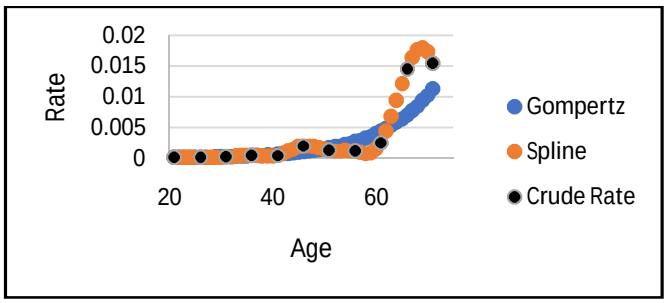

Figure 5(a): Fitted Cancer Incidence Rate – Male (FH)

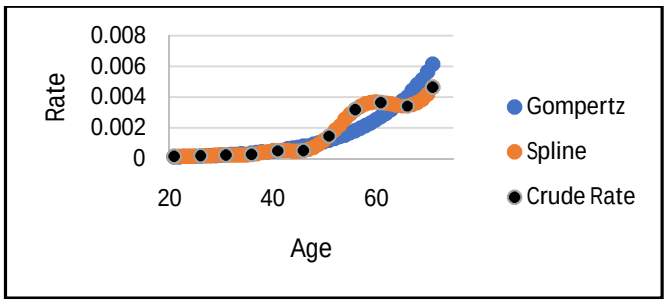

Figure 5(b): Fitted Cancer Incidence Rate – Male (NFH)

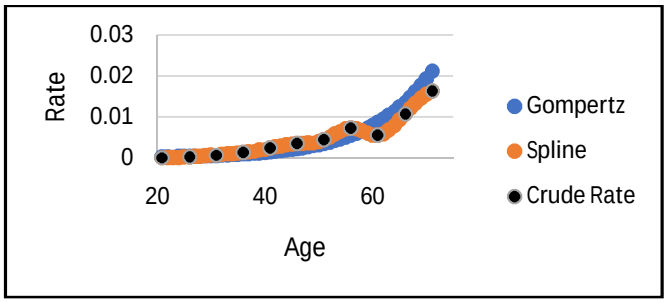

Figure 6(a): Fitted Cancer Incidence Rate – Female (FH)

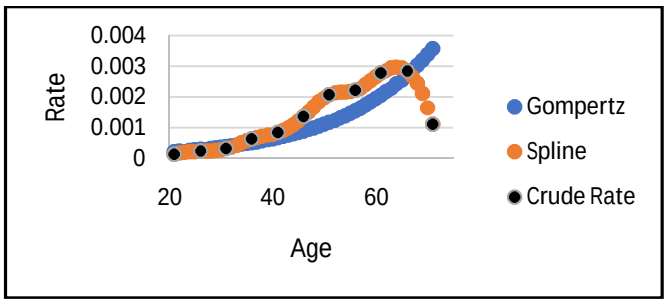

Figure 6(b): Fitted Cancer Incidence Rate – Female (NFH)

The spline graduated line shows a better fitting for both gender categories and both groups, with and without family history of cancer as the incidence rate across ages has of a non-linear curve pattern. As depicted in Figure 5(a) and 5(b), the crude cancer incidence rates at higher ages increase substantially for people with and without family history of cancer for male category. In contrast, for female category, the crude cancer incidence rates increase substantially at age 70 only for people with family history of cancer but not for people without family history of cancer (refer to Figure 6(a) and 6(b)). This indicates that the use of family history information in cancer insurance pricing may affect the female category more than the male category due to this difference.

4.3 Cost of Insurance with Family History Information

In this section, the transition probability for each event, i.e. healthy-to-healthy (00), healthy-to-cancer (01) and healthy-to-death (02) is estimated and used to calculate the risk-based pricing of cancer insurance coverage. The results of insurance Net Single Premium at selected purchase age is determined for both groups, with and without family history of cancer and the difference in amount is compared and shown in Table 2(a) and Table 2(b) for male and female category respectively.

Based on our results in Table 2(a), the net single premium for people with family history of cancer is considerably higher than people without family history of cancer. For male category, the difference in premium slowly increases as age of purchase increases, from 35.93% for purchase age of 25 years old to 47.79% for purchase age of 55 years old except for purchase age of 50 years old where it drops slightly to 33.92%. This can be explained by the large increase of cancer incidence rate for people without family history of cancer for ages between 50 to 60 years old as depicted by Figure 5(b).

Similarly, for female category, the net single premium for people with family history of cancer is higher but the magnitude is substantially higher than the male category. As shown in Table 2(b), the percentage increase in premium for people with family history of cancer is between 127.58% for purchase age of 25 years old to 149.42% for purchase age of 55 years old. This is supported by our cancer incidence rates estimation for female category as illustrated graphically in Figure 5(a) and 5(b), where the incidence rate increases at higher ages only for females with family history of cancer but drops significantly for females without family history of cancer.

Table 2(a): Difference of Net Single Premium for Males

Age	Family History (RM)	No Family History (RM)	Difference (RM)	% increase
25	1837.47	1351.73	485.74	35.93
30	2177.22	1575.41	601.81	38.20
35	2550.25	1830.36	719.88	39.33
40	2970.38	2106.44	863.94	41.01
45	3367.09	2380.14	986.94	41.47
50	3570.52	2666.25	904.27	33.92
55	3969.76	2686.12	1283.65	47.79

Table 2(b): Difference of Net Single Premium for Females

Age	Family History (RM)	No Family History (RM)	Difference (RM)	% increase
25	2895.52	1272.30	1623.22	127.58
30	3439.84	1482.37	1957.48	132.05
35	3971.83	1699.25	2272.58	133.74
40	4460.90	1885.13	2575.77	136.64
45	4779.05	2040.13	2738.92	134.25
50	4956.75	2064.89	2891.85	140.05
55	4789.66	1920.29	2869.37	149.42

Thus, our results indicate that family history is an important risk factor in calculating the premium for cancer insurance coverage especially for female category. If family history information is unavailable, or a standard premium is charged for everyone without taking into account the risk factor of family history of the disease, individuals without family history of cancer will be highly subsidizing the individuals with family history of cancer.

5. CONCLUSION

Cancer disease has been a leading cause of death in Malaysia in the past and causing financial burden to patients diagnosed with the disease due to the expensive cancer related treatments. The heterogeneous sources of cancer morbidity risk within a population may be due to different life style, food consumption, and the existence of family history of the disease. This research aims to evaluate the implication of using a family history information as a risk factor on cancer incidence rate and the cost of cancer insurance.

There is an extensive ongoing argument on whether genetic testing should be used as a risk factor in insurance pricing which contains a detailed result on family inheritance of cancer. Some countries ban the use of genetic testing and some have budget constraint due to the expensive cost of this testing. Following these constraints, family history is the best information available to assess on any chance that a potential insured is carrying a high risk of a particular disease.

Risk-based pricing is crucial to ensure an accurate and fair premium is charged to insurance policyholders.

Based on our results, family history significantly affects the net single premium of cancer insurance coverage especially for female category. The fair premium cost for people with family history of cancer is more than double in comparison to people without family history of cancer for female category at all purchase ages. Hence, the presence of additional information of family history improves the ability to price an insurance more accurately which consequently reduce the problem of mispricing and adverse selection in the insurance market.

ACKNOWLEDGEMENT

This study is funded by UniversitiTeknologi MARA through the Lestari grant under code 600-IRMI 5/3/LESTARI (031/2019). The main author would like to acknowledge the financial support receive from the Ministry of Education Postdoctoral Scholarship during her postdoctoral study at the University of Waterloo, Canada.

REFERENCES

1. World Health Organization, **"Global Cancer Rates could Increase by 50% to 15 Million by 2020,"** ***News Release***, 2003. [Online]. Available: https://www.who.int/mediacentre/news/releases/2003/pr27/en/. [Accessed: 11-Feb-2020].
2. Y. Li, J. Schoufour, D. D. Wang, K. Dhana, A. Pan, X. Liu, … F. B. Hu, **"Healthy Lifestyle and Life Expectancy Free of Cancer, Cardiovascular Disease, and Type 2 Diabetes: Prospective Cohort Study,"***BMJ*, vol. 368, January, pp. 1–10, 2020.
3. A. Macdonald and F. Yu, **"The Impact of Genetic Information on the Insurance Industry: Conclusions from the 'Bottom-up' Modelling Programme,"***ASTIN Bulletin*, vol. 41, no. 2, pp. 343–376, 2011.
4. R. C. W. B. Howard, **"Genetic Testing Model for CI : If Underwriters of Individual Critical Illness Insurance Had No Access to Known Results of Genetic Tests,"** Canadian Institute of Actuaries Research Report, 2016.
5. B. Lu, **"Some New Actuarial Models of the Insurance Implications of Genetic Testing for Breast and Ovarian Cancer,"** Heriot-Watt University Ph.D Thesis, 2007.
6. C. Y. Seng and M. N. Isa, **"Guideline of the Malaysian Medical Council: Medical Genetics and Genetic Services,"** 2006.
7. C. Wekwete, **"Genetics and Critical Illness Insurance Underwriting: Models for Breast Cancer and Ovarian Cancer and for Coronary Heart Disease and Stroke,"** Heriot-Watt University Ph.D Thesis, 2002.
8. M. B. Moghadam, **"Extended Risk Classification in Insurance Industry,"***Quality and Quantity*, vol. 47, pp. 1385–1396, 2013. https://doi.org/10.1007/s11135-011-9596-9
9. A. S. Macdonald and J. Lemaire, **"Genetics, Family History and Insurance Underwriting: an Expensive Combination?,"** ASTIN Colloquia, 2003.
10. P. Elko, R. Pravasta, T. William, G. Wang, and E. R. Kaburuan, **"Study of Indonesia Vehicle Insurance for New Scheme Applied Use Based Insurance Model,"***International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 6, pp. 2726–2729, 2019. https://doi.org/10.30534/ijatcse/2019/07862019
11. National Cancer Registry; National Cancer Institute; Ministry of Health Malaysia, ***Malaysian Study on Cancer Survival (MySCan)***, vol. 4. 2018.
12. **"Cutting Cancer Treatment Costs,"***The Star Online*, 2019. [Online]. Available: https://www.thestar.com.my/news/nation/2019/03/24/cutting-cancer-treatment-costs. [Accessed: 11-Feb-2020].
13. S. K. Veettil, K. G. Lim, N. Chaiyakunapruk, S. M. Ching, and M. R. Abu Hassan, **"Colorectal Cancer in Malaysia: Its Burden and Implications for a Multiethnic Country,"***Asian Journal of Surgery*, vol. 40, pp. 481–489, 2017.
14. S. Surendran, **"An Opportune Time to Help the M40,"***The Edge Markets*, 2020. [Online]. Available: https://www.theedgemarkets.com/article/cover-story-opportune-time-help-m40.[Accessed: 18-May-2020].
15. S. J. A. Ibrahim and M. Thangamani, **"Innovative Drug and Disease Prediction with Dimensionality Reduction and Intelligence Based Random Walk Methods,"***International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 4, pp. 1668–1673, 2019. https://doi.org/10.30534/ijatcse/2019/93842019
16. C. A. Alexander and L. Wang, **"Big Data Analytics in Heart Attack Prediction,"***Journal of Nursing & Care*, vol. 06, no. 02, pp. 1–9, 2017.
17. N. Malik, V. B. Bharat, S. P. Tiwari, and J. Singla, **"Study of Detection of Various Types of Cancers by using Deep Learning: A Survey,"***International Journalof Advanced Trends in Computer Science and Engineering*, vol. 8, no. 4, pp. 1228–1233, 2019. https://doi.org/10.30534/ijatcse/2019/31842019
18. A. D. Wilkie, **"Mutuality and Solidarity: Assessing Risks and Sharing Losses,"***British Actuarial Journal*, vol. 3, no. 5, pp. 985–996, 1997.
19. E. Hock Gui, B. Lu, A. Macdonald, H. Waters, and C. Wekwete, **"The Genetics of Breast and Ovarian Cancer III: A New Model of Family History with Insurance Applications,"***Scandinavian Actuarial Journal*, vol. 2006, no. 6, pp. 338–367, 2006.
20. D. R. Cox, **"Regression Models and Life-Tables,"***Journal of the Royal Statistical Society*, vol. 34, no. 2, pp. 187–220, 1972.
21. J. Fox and S. Weisberg, **"Cox Proportional-Hazards Regression for Survival Data in R,"** in An R Companion to Applied Regression, Third Edition, Appendix, California: Sage Publications, 2018, pp. 1–18.

22. I. J. Myung, **"Tutorial on Maximum Likelihood Estimation,"***Journal of Mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.
23. D. O. Forfar, J. J. McCutcheon, and A. D. Wilkie, **"On Graduation By Mathematical Formula,"***Transactions of the Faculty of Actuaries*, vol. 115, no. 1, pp. 1–149, 1988.
24. E. Dodd, J. J. Forster, J. Bijak, and P. W. F. Smith, **"Smoothing Mortality Data: the English Life Tables, 2010–2012,"***Journal of the Royal Statistical Society. Series A: Statistics in Society*, vol. 181, no. 3, pp. 717–735, 2018.
25. A. Perperoglou, W. Sauerbrei, M. Abrahamowicz, and M. Schmid, **"A Review of Spline Function Procedures in R,"***BMC Medical Research Methodology*, vol. 19, no. 46, pp. 1–16, 2019.
26. A. S. Macdonald, H. R. Waters, and C. T. Wekwete, **"The Genetics of Breast and Ovarian Cancer I: A Model of Family History,"** *Scandinavian Actuarial Journal*, vol. 2003, no. 1, pp. 1–27, 2003. https://doi.org/10.1080/03461230308486
27. Azizah AM., Hashimah B., Nirmal K., Siti Zubaidah AR., Puteri NA., Nabihah A., … Azlina AA., ***Malaysia National Cancer Registry Report 2012-2016.*** 2019.
28. P. Gatenby, **"Long Term Care,"** Staple Inn Actuarial Society, 1991.