International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 4, April - 2022, Pages:132-141

The Impact of Health Insurance on the Services Provided to the Citizen: A Parametric Survival Models

Salma Abass Basher Abass[1] and Altaiyb Omer Ahmed Mohmmed[2]

1,2 Sudan university of science and technology, college of science, statistics department

1 corresponding author's email: Nazar9090215@gmail.com

Abstract : *This study aimed to compare the parametric survival models such as exponential model, Weibull model, Log-logistic model, lognormal model and Gompertz model to determine the best fits model to estimate and evalute the service provided from health insurance to the citizen in khartoum region subscriptions that benefits from service provided from hospital or pharmacies. The Models were estimated by using maximum likelihood method, and Log-likelihood and Akaike Information Criterion (AIC), paysian criterion (BIC) were used as selection criterion methods. The results of AIC and Log-likelihood show that Log-logistics regression model best fit the the servise provided from health insurance to the citizen. The results of exponential regression model revealed that revenues and benefits are prognostic factors affecting the the service that provided from health insurance to the citizen. Therefore, it is recommended that researchers of the health care unit consider this model (exponential regression model) in their researches concerning other field.*

Keywords: health insurance, maximum likelihood, prognostic factors, parametric survival models

Methodology and material

Data were collected from the Health Insurance corporate Khartoum state for the targeted residents of the state in various localities, as well as the percentage of insurance coverage, frequency of movement of patients, the total cost and benefits paid to the institution providing the service. And to study the movement of patients and visitors to government hospitals and private clinics, MINTAB-SPSS- STATA were used for data analysis.

Introduction

Universal health coverage (UHC), which means that everyone can access sufficient quality health care services, including promotive, preventive, curative, rehabilitative, and palliative services without any financial difficulties, is a widely used concept, especially in low- and middleincome countries (LMIC) [1]. The World Health Organization is guiding LMIC to develop a health financing system to achieve and maintain UHC, where the national HI system has been promoted as a vital health financing strategy to expand pooled funds for equitable financing of health care [2].

Health insurance is one of the types of insurance against the risks of health conditions for individuals, and includes the costs of examination, diagnosis, treatment, and psychological and physical support [3]. It may also include coverage for his interruption from work for a certain period or his permanent disability. It is one way to deliver health care to individuals and groups. The philosophy of health insurance is based on the principle of aggregation of risks, which means collecting the risks of disease that affect population or a particular group, and sharing them equally among individuals, by collecting the necessary funds to treat this combined risk equally, and then distributing them to individuals according to their need for treatment, which leads to alleviating Burdens and costs incurred when treating sick conditions to which the insured are exposed, and ensuring access to health care for all those in need [5].

In this aspect, some statistics models have been used, which are usually defined on the positive (non-negative) period {$0, \infty$}. These models are used to analyze the relative risks that occur on the types of insurances, and here, we single out health insurance as one type of insurance that interests the citizen and is one of his priorities, and these are distinguished for a certain amount of the money paid by all the individuals participating in the insurance [2].

The value of the monthly subscription premium for the family is determined based on the consumption of the service. The services provided to subscribers are also priced by calculating the costs of medical inputs, economic changes, the rise in the exchange rate, the hard currency, and the customs of imported goods that are included in the health integration system. The increase in the prices of medical and pharmaceutical service inputs has led to Increasing costs for medicines and, consequently, all other medical services accessories, which prompted the authority to renew the prices of the service provided through the contracted institutions, in addition to the increase in visitors to medical services in light of the stability of revenues (the value of the subscription for the family), which led to an increase in the cost of the service provided. It is expected that the authority will face some problems related to fulfilling its obligations towards subscribers and service providers [3].

The different types of health insurance

The various policy types and health plan names will prepare you for evaluating your options when you're ready to enroll in a new plan. The more familiar you are with the different insurance plan types, the better equipped you'll be to pick one to fit your company's budget and needs [2].

Preferred provider organizations (PPOs)

With a (PPO) plan, employees are encouraged to use a network of preferred doctors and hospitals to receive their medical needs at a negotiated or discounted rate. Employees generally aren't required to select a primary care physician (PCP) and have the choice to see any doctors within their network [2].

Health maintenance organizations (HMOs)

Employees with an (HMO) generally have lower out-of-pocket expenses but less flexibility in their choice of physicians or hospitals than other plans. An HMO usually requires employees to choose a PCP as part of their plan and employees need to obtain a referral from their PCP to see a specialist [1].

Point of service (POS) plans

A (POS) group health plan combines features of an HMO and a PPO plan. Just like an HMO, POS plans may require employees to choose a PCP from the plan's network providers. Generally, services rendered by the PCP aren't subject to the policy's deductible [3].

Exclusive provider organization (EPO) plans

EPO plans are similar to HMOs because they have a network of physicians their members are required to use except in emergencies. Members have a PCP who provide referrals to in-network specialists and members are also responsible for small co-payments and potentially a deductible [1].

Indemnity plans

Indemnity health plans are known as fee-for-service plans. With indemnity plans, the insurance company pays a predetermined percentage of the reasonable and customary charges, or the average fee within a geographic area, for a given service, and the insured pays the rest [2].

Health savings accounts (HSAs)

An HSA is a tax-advantaged savings account used in conjunction with an HSA-compatible high deductible health plan (HDHP) to pay for qualifying medical expenses. Though HSAs can be attached to group health insurance, employers can contribute to the account whether they offer a group policy or not, and the account goes with the employee when they leave the company. However, you can only contribute to an HSA if you have a HDHP [1].

Parametric Survival models

A parametric survival model time is supposed to follow a certain distribution, which its probability density function can be represented by unknown parameters. Weibull, Exponential, Gompertz, log logistic, lognormal and gamma distributions are widely used[1,8].

Survival analysis functions

Let be (T) represents a (non-negative) continuous random variable and represents the waiting time (survival time) from the moment of observing the element until the occurrence of the event to be studied[4]. The probability distribution function of the survival function of the random variable (T) is the probability distribution function, which is defined as:

$$F(t) = p(T < t), t \geq 0 \quad \text{..............................(1)}$$

This function gives the probability that the element's survival time in the population does not exceed from the orginal time, or the probability that the element get event. The survival time can be also described by the survival function as shown in formula (2) below:

$$S(t) = P\{T \geq t\} = 1 - F(t), \quad t \geq 0 \quad \text{..........................(2)}$$

And this function gives two possibilities [11] either: the survival time exceeds the time moment, or the faliur time occurs in the time period.

The survival function can be expressed as a density function to describe the survival time, which is usually derived from the distribution function:

$$f(t) = F'(t) = -S'(t) \quad \text{...(3)}$$

Since the random variable (T) a continuous random variable is defined in the interval { $0, \infty$ }, equation (3) can be rewritten as:

$$F(t) = \int_0^1 f(u)du \quad \text{..(4)}$$

$$S(t) = \int_1^\infty f(u)du \quad \text{..(5)}$$

The survival function is usually used in survival analysis, although the probability distribution function is more commonly used in other fields of statistics [7]. It is noted that the survival function has the following properties [4]:

It's a non-increasing function for all:

$$S_{\lim_{t \to \infty}}(t) \geq S(t + \alpha), \forall \alpha > 0 \quad \text{..............................(6)}$$

Where equation (6) can be rewritten as:

$$S(t) = 1 - F(t) \quad \text{...(7)}$$

From equation (7) we get:

$$\frac{dS(t)}{dt} = -f(t) \leq 0, \qquad t \geq 0 \quad \text{...................................(8)}$$

At the start of the study, i.e. time zero, all the elements are still alive, and this is expressed as:

$$S(0) = 1 \quad \text{...(9)}$$

At the end of time, i.e. infinty time, the survival function becomes:

$$\lim_{1 \to \infty} S(t) = 0 \quad \text{..(10)}$$

Hazard Function:

The hazard function, denoted by h(t), is given by the formula [6]: h(t) equals the limit, as Δt approaches zero, of a probability statement about survival, divided by Δt, where Δt denotes a small interval of time. This mathematical formula is difficult to explain in practical terms. usually given by formulas [8]:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta T \backslash T > t)}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{P\left(t < T \le t + \Delta t, T < t\right)}{P(T > t).\Delta t} = \frac{f(t)}{s(t)} \quad \ldots\ldots\ldots\ldots(11)$$

Considering the $-f(t)$ is the derivative from, the hazard function it can be rewritten as:

$$h(t) = -\frac{d}{dt}\ln S(t) \quad \ldots\ldots\ldots\ldots(12)$$

By integrating both sides of equation (12) for the time interval {$0, t$}, we get:

$$\int_0^1 h(u)du = -InS(t) \quad \ldots\ldots\ldots\ldots(13)$$

$$\Rightarrow S(t) = e^{-\int_0^1 h(u)du} \quad \ldots\ldots\ldots\ldots(14)$$

$$H(t) = \int_0^1 h(u)du \quad \ldots\ldots\ldots\ldots(15)$$

equation (15) is called the Cumulative Hazard Function, which represents the hazard from the start time 0 up to time (t) [8].

Expected survival time

Let be μ is average survival time and given by:

$$\mu = E(T) = \int_0^\infty tf(t)dt \quad \ldots\ldots\ldots\ldots(16)$$

And since $-f(t)$ it is a derivative from $s(t)$, and based on the equations (9) (10), by take the integration to get the expected survival time [11]:

$$E(T) = \int_0^\infty S(t)dt \quad \ldots\ldots\ldots\ldots(17)$$

The expected value for any element that has survived until to get event in time (t_0) is given by:

$$E\left[T-t_0 \backslash T \geq t_0\right] \quad \ldots(18)$$

And then:

$$E\left[T-t_0 \backslash T \geq t_0\right]=\int_0^{\infty} t \; dp\left(T-t_0 \leq t \backslash T>t_0\right)=\int_0^{\infty} t \; dp\left(T \leq t_{0^+} t \backslash T>t_0\right) \quad \ldots(19)$$

Thus, the remainder of the distribution function for the survival time is given by:

$$p\left(T \leq t_0+t \backslash T>t_0\right)=\frac{p\left(t_0<T \leq t_0+t\right)}{p\left(T>t_0\right)}=\frac{F\left(t_0+t\right)-F+\left(t_0\right)}{S\left(t_0\right)} \quad \ldots(20)$$

And the density function is:

$$\frac{d}{dt} \frac{F\left(t_0+t\right)-F\left(t_0\right)}{S\left(t_0\right)}=f \frac{\left(t_0+t\right)}{S\left(t_0\right)} \quad \ldots(21)$$

The expected survival time is obtained from [5]:

$$E\left[T-t_0 \backslash T \geq t_0\right]=\frac{1}{S\left(t_0\right)} \int_0^{\infty} t \; f\left(t+t_0\right) dt \quad \ldots(22)$$

Any distribution defined in the interval { $t \in (0, \infty$ } can be used as a distribution of the survival time, knowing that we can use some of the distributions defined in the interval { $y \in (-\infty, +\infty)$ }, if we assume that { $t = e^y$ } and by setting { $y = \ln t$ }. Since Y is a random normal variable defined in the interval { $y \in (-\infty, +\infty)$ }, the variable { $t = e^y$ } it is subject to the logarithmic normal distribution defined in the interval { $t \in (0, \infty$ }.

The exponential, logistic, weibull and gompertz distributions are considered one of the most important and most famous probability distributions for the survival time, in addition to the diversity of trends in hazard functions among the models between fixed as in exponential and decreasing as in logistic, increasing as in Gompertz, and multi-directional as in weibull and lognormal [13].

Table (1): the most important survival distributions:

Probability distribution	Survivor function	Parameter
Exponential	$e^{-\lambda t}$	$\lambda > 0$
Lognormal	$1-\phi = \frac{\ln t - \mu}{\sigma}$	$\mu \in R, \sigma > 0$
Logistic	$\{1+(\frac{t}{\alpha})^{\beta}\}^{-1}$	$\alpha > 0, \beta > 0$
Weibull	$e^{-(\lambda t)\gamma}$	$\lambda > 0, \gamma > 0$

Gompertz	$e^{\lambda(1-e^{\vartheta t})}$	$\lambda > 0, \vartheta \in R$

Source: machin 2008

Exponential Distribution

The exponential distribution is one of the simplest distributions representing the survival time, and it has one coefficient, but it has many applications in all aspects of life [14]. that the exponential distribution has a probability density function that gives:

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, t \geq 0 \\ 0, \quad t < 0 \end{cases} \qquad (23)$$

Note that ($\lambda > 0$) is the parameter of the distribution.

A high λ value shows high risk and limited survival; a low λ value shows low risk and long survival. The distribution is also referred to as the unit exponential when $\lambda = 1$ [11].

The probability distribution function is given by the formula:

$$f(t) = \begin{cases} 1 - e^{-\lambda t}, t \geq 0 \\ 0, \quad , t < 0 \end{cases} \qquad (24)$$

And the moments in an exponential distribution are given by:

$$ET\,2 = \int_o^\infty t^r \lambda e^{-\lambda t} .dt \Rightarrow ET^r = \frac{r!}{\lambda^r} \quad , r = 1, 2, \ldots \qquad (25)$$

to find the measures of central tendency, substituted values in equation (25) to get the arithmetic mean [10].

by substituting r = 1 into equation (25), we get the expected value as in equation (26):

$$ET = \frac{1}{\lambda} \qquad (26)$$

to find the median, we solve the equation $F(t) = \frac{1}{2}$ then the median is equal to $\frac{\ln(2)}{\lambda}$, also to find the variance as one of the measures of dispersion, we get it as equation (27) [4]:

$$v(t) = E\left(T - ET\right)^2 = \frac{1}{\lambda^2} \qquad (27)$$

To find the skewness, it can be obtained as

$$sk = \frac{E(T - ET)^2}{\left[V(T)\right]^3} = 2 \qquad (28)$$

That is, implies the distribution is asymmetric with the mean and is skewed to the right.

Exponential Survival function

The survival function for an exponential distribution is given as:

$$S\left(t\right)=e^{-\lambda t},t\geq 0 \qquad \text{………………………………………………………………………(29)}$$

That is means $e^{-\lambda t}$, it represents the probability of any event will survive in the interval { $0,t$ } at least.

Exponential hazard function

Recalled to equation (10), we find that $h\left(t\right)=\lambda$, this is implies the exponential hazard function is fixed for each 0≤t, and it represents the rate of event occurrence in time. Therefore, it is obvious that we expect that the exponential distribution does not agree with cases in which the rate of hazard varies with time [7], such as the failure of the machine, whose probability of failure increases with time. But in some cases, the death rate can be constant, for example, the death rate (shatter) of glassware is considered constant, since the rate of death (shattered) of the glassware is fixed, that is, it has nothing to do with time. This explains why the exponential distribution is memoryless, which is expressed by:

$$P(T>s+t\setminus T>s)=P(T>t);\forall s,t\geq 0 \qquad \text{…………………………………………..(30)}$$

This is proves that any event has a constant with time, so the time expected until the event occurs for the first time is an exponential distribution also.

log-logistic distribution

In probability and statistics, the log-logistic distribution is a continuous probability distribution for a non-negative random variable [14]. It is used in survival analysis as a parametric model for events whose rate increases initially and decreases later, as, for example, mortality rate from cancer following diagnosis or treatment.

Probability Density function

The probability density function of log-logistic distribution it given as:

$$f(t)=\frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]^{2}},t>0,\alpha>0,\beta>0 \qquad \text{…………………………………………..(31)}$$

Survival function

$$s(t)=\left[1+(t/\alpha)^{\beta}\right]^{-1},t>0,a>0,\beta>0 \qquad \text{………………………………………………(32)}$$

Hazard function

$$h(t)=\frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{1+\left(\frac{t}{\alpha}\right)^{\beta}},t>0,\alpha>0,\beta>0 \qquad \text{………………………………………………(33)}$$

Expected value

$$E(T)=\frac{\alpha b}{\sin(b)}, \quad b=\frac{\pi}{\beta}; \quad \alpha>0, \beta>1 \qquad \text{..........(34)}$$

Variance

$$V(T)=\alpha^2\left(\frac{2b}{\sin 2b}-\frac{b^2}{\sin^2 b}\right), \qquad \alpha>0, \beta>2 \qquad \text{..........(35)}$$

Gompertz distribution

In probability and statistics, the Gompertz distribution is a continuous probability distribution, named after Benjamin Gompertz. The Gompertz distribution is often applied to describe the distribution of adult lifespans by demographers [preston,2001]and actuaries. [willemse & koppelaar,2000] also considered the Gompertz distribution for the analysis of survival. More recently, computer scientists have also started to model the failure rates of computer code by the Gompertz distribution.[13].

Density Distribution Function

The probability density function of Gompertz distribution is given as

$$f(t)=abe^{bt}e^{a}e^{-ae^{bt}}, t\geq 0, a>0, b>0 \qquad \text{..........(36)}$$

the distribution function

$$F(T)=1-e^{-ae^{bt}}, \qquad t>0 \qquad \text{..........(37)}$$

Selection Criterion

One of these criteria is the information criterion of Akaike (AIC), the Baysian Information Criterion (BIC) and the Cox-Snell Information Criterion (CSIC), the latter of which is a graphic rather than a mathematical criterion, many of the criteria used to choose the best model from diff erent models deal with the same data for prediction in the future.

AIC: Comparisons may also be made on the basis of statistics between a variety of potential models which do not necessarily need to be nested [9].

$$AIC=-2\log likelihood+2(P+K)$$

Where P is the number of parameters, and K is the number of (excluding constant) coeffi cients in the model. For P=1, for P=2, for Weibull and Gompertz, for the exponential. Th e smaller the value of this statistic, the better the model, the better this statistic is known

as Akaike’s knowledge criterion.

Application with dataset

analyse survival probabilities of the patients. compared six parametric survival models; Exponential, Weibull, log-normal, log-logistic, Gompertz and revenues and benefits models and their performance were assessed using Akaike information Criterion (AIC) and Bayesian Information Criterion (BIC) goodness of fit criteria to determine the best model.

Table (2): the targeted movement Statistics during the study period

Year	Target population	Entry within year	Input-to-target ratio
2016	1317101	77824	5.91
2017	1351514	84815	6.28
2018	1386357	109932	7.93
2019	1421626	103059	7.25
2020	1457305	21904	1.50

The target population and real entry in same year were conduct to shows how so far the gap between years, also ratio between them calculated, in the last two year 2020 with lower ratio.

Table (3): Descriptive statistics

	Sum	Mean	Std. deviation	Std. Error of Mean
revenues	4292245021.95	85844900.44	201105187.23	28440568.32
benefits	3863242568.94	77264851.38	122564236.90	17333200.61

Descriptive statistics for the variables studied and their characteristics for the all years to shows the target population to server him.

The maximum likelihood estimates are calculated directly by using by maximizing the likelihood function. We have obtained the values of Log-Likelihood, Akaike information criterion (AIC), Bayesian information criterion (BIC), Corrected Akaike information criterion (CAIC) and Hannan-Quinn information criterion (HQIC).

Table (4): Best criteria model

Model	-LL	AIC	CAIC	BIC	HQIC
Gompertz	171.516	336.5214	336.8745	342.5423	339.1230
Exponential	147.362	312.4851	313.0123	319.6423	317.1029
Gamma	181.418	362.5170	363.1452	365.4120	363.9842
Weibull	199.835	384.1231	385.2145	388.0213	386.7542
Logistic	175.689	340.7451	340.9458	345.8461	342.8109

Table (5): Goodness of fit statistics

Model	Ks	W	A^2
Gompertz	0.0516	0.0214	0.1745
Exponential	0.0362	0.0851	0.2123
Gamma	0.1418	0.2170	0.4402
Weibull	0.0835	0.1231	0.9155
Logistic	0.0689	0.0451	1.1458

To compare the goodness-of-fit of the all distribution of each other we have calculate the value of Kolmogorov-Simnorov (KS), Anderson-Darling (W) and the Cramer-Von Mises (A^2). It is show that the LG distribution has the minimum value of the test statistic and thus we conclude that the LG distribution gets a better fit and more consistent and reliable results from others taken for comparison.

Conclusion

To find out the best model, we have developed many distribution studied such as Gompertz distribution, Exponential distribution, Gamma distribution, Weibull distribution, Logistic distribution . then we have provided the statistical properties of the best proposed model. The application illustrates that the proposed model provides a consistently better fit and more flexible than other underling models. And exponential model was best model according to the goodness of fit criteria. We expect that this model will contribute to the field of the best fits model to estimate and evalute the servise provided from health insurance to the citizen in khartoum region subscriptions.

Refreneces

[1] S. Prinja, P. Bahuguna, I. Gupta, S. Chowdhury, and M. Trivedi, “Role of insurance in determining utilization of healthcare and financial risk protection in India,” PLoS One, vol. 14, no. 2, article e0211793, 2019.

[2] R. Pokharel and P. R. Silwal, "Social health insurance in Nepal: a health system departure toward the universal health coverage,"The International Journal of Health Planning and Management, vol. 33, no. 3, pp. 573–580, 2018.

[3] S. Prinja, G. Kaur, R. Gupta, S. K. Rana, and A. K. Aggarwal,"Out-of-pocket expenditure for health care: district level estimates for Haryana state in India," The International Journal of Health Planning and Management, vol. 34, no. 1, pp. 277–293, 2019.

[4] Kleinbaum, David G, and Mitchel Klein (2012). Survival AnalysisA Self-Learning Tex. Third Edition. New York: Springer Science + Business Media.

[5] Collett, David. (2003). Modelling Survival Data in Medical Research. Second Edition. London: Chapman & Hall.

[6] Lee, Elisa T, and John Wenyu Wang.(2003). "Some Well-Known Parametric Survival Distributions and Their Applications", Statistical Methods for Survival Data Analysis. Third Edition.Canada. John Wiley & Sons, Inc., Hoboken, New Jersey.

[7] George, Brandon, Samantha Seals, and Inmaculada Aban. (2014). Survival analysis and regression models.J Nucl Cardiol, 21(4):686–694.

[8] Abdelaal, Medhat Mohamed Ahmed, and Sally Hossam Eldin Ahmed Zakria. (2015). Modeling Survival Data by Using Cox Regression Model. American Journal of Theoretical and Applied Statistics,4(6):504-512.

[9] Lawless, Jerald F. (2003). Statistical Models and Methods for Lifetime Data. Second Edition. New York: John Wiley & Sons.

[10] Pourhoseingholi, Mohamad Amin, Ebrahim Hajizadeh, Bijan Moghimi Dehkordi, Azadeh Safaee, Alireza Abadi, and Mohammad Reza Zali.(2007).Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma. Asian Pacifi c J Cancer Prev,8: 412-416.

[11] Klein, John P, and Melvin L Moeschberger.(1997).Survival Analysis: Techniques for Censored and Truncated Data .New York: Springer-verlag.

[12] Preston, Samuel H.; Heuveline, Patrick; Guillot, Michel (2001). *Demography: measuring and modeling population processes*. Oxford: Blackwell.

[13] Willemse, W. J.; Koppelaar, H. (2000). "Knowledge elicitation of Gompertz' law of mortality". *Scandinavian Actuarial Journal*. **2000** (2): 168–179. doi:10.1080/034612300750066845.

[14] Ohishi, K.; Okamura, H.; Dohi, T. (2009). "Gompertz software reliability model: estimation algorithm and empirical validation". *Journal of Systems and Software*. **82** (3): 535–543. doi:10.1016/j.jss.2008.11.840