World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

The Cox regression and Kaplan-Meier for time-to-event of survival data patients with renal failure

Ayat Mubarak Karamalla Elamin * and Altaiyb Omer Ahmed Mohmmed

Sudan University of Science and Technology, College of Science, Sudan.

World Journal of Advanced Engineering Technology and Sciences, 2023, 08(01), 097–109

Publication history: Received on 28 November 2022; revised on 15 January 2023; accepted on 18 January 2023

Article DOI: https://doi.org/10.30574/wjaets.2023.8.1.0183

Abstract

Background: renal failure disease usually occurs when the blood supply to the kidneys is suddenly interrupted or when the kidneys become overloaded with toxins.

Objectives: in this article two main survival models were used, Cox regression and Kaplan-Meier to estimate the median patient survival time for conditions causing renal failure and to comparison of the survival rates of people with illnesses causing renal failure.

Methods: The research community is made up of individuals who have been diagnosed with renal failure, and the data were gathered from the patient records at the Police Hospital (Khartoum - Burri). All individuals with renal failure who were tracked down and given the diagnosis were included in the process of thorough inventory. The computations were done using some statistical Software (SPSS, STATA), with level of significance 0.05.

Results: some variables like other disease were causes RF (diabetes, heart disease, osteoporosis, hepatitis, and growth retardation) associated with renal failure, Cox regression was used and the basic variables shown that 65.7% of the independent variables (duration of disease, housing, diabetes and hepatitis infection) determine the survival time of the estimated model.

Conclusions: Kaplan–Meier and Cox regression methods both are used in clinical and epidemiological research. The Cox regression analysis is based on estimating the HR associated with a specific risk factor or predictor for a given endpoint. The standard Cox regression method allows for an investigation of the effect of one or more variables (covariates) on the "time-to-first-event" analysis. An assessment of proportional hazards is a prerequisite to fitting a Cox regression model.

Keywords: Hemodialysis; Survival Analysis; Parametric; Proportional Hazards (PH); Modeling

1. Introduction

The survival analysis is a branch of statistics to analyze the expected time until the occurrence of a single event, such as death in living organisms and failure in mechanical systems (Kleinbaum 2005). The time before and after the event under study, such as the event of death, birth, or follow-up to the (disease classification), is what makes this conclusion (mantel 1966).

The major aim of survival analysis is the modeling and analysis of 'time-to-event' data; that is, data that have as an end point the time when an event occurs. In this respect, events are not limited to death but may include all kinds of 'positive'

* Corresponding author: Ayat Mubarak Karamalla Elamin

or 'negative' events like myocardial infarction, recovery of renal function, first renal transplant, graft failure, or time to discharge from hospital (Kitty et al, 2008).

The Kaplan-Meier method estimates the survival function, which is the probability of "surviving" beyond a certain time point. The Kaplan-Meier procedure (Kaplan & Meier, 1958) is used to calculate the survival rate from the survival function. It involves estimating the probability of surviving for a specified length of time. The advantage of using Kaplan-Meier curves is that they are non-parametric, where no assumptions are made on the distribution of survival times (Daniel, 2005). The survival function is the number of individuals with survival time, which is at least t time periods divided by the number of individuals in the study.

The Kaplan–Meier (KM) method explores the survival of a population under investigation and/or tests differences in the crude cumulative survival between exposure groups, with a graphical representation of the endpoint occurrence as a function of time (Goel et al 2010).

The Cox regression model is probably the most popular regression technique for regression analysis of survival data (Hassan et al 2014).

2. Material and methods

This is a retrospective cohort study, which evaluated the data of 130 patients on renal failure, over a period of 20 years. All patients were followed retrospectively until death or censorship (end of follow-up, loss of follow-up, transfer or kidney transplant). Of the 130 patients, 66(51%) were censored and 64(49%) failed (dead). Data were obtained from medical records of patients undergoing treatment and from the computerized record system of the police hospital in Burri count Khartoum region, where the study was conducted. For this study, we included consecutive patients with ESRD at any time during the study follow-up, with a minimum treatment time of one month, age ≥ 18 years and with all available baseline data. To be diagnosed with RF, and the diagnosis of disease was based on clinical, laboratory and radiological characteristics.

2.1. Cox Regression concept

Cox regression (or proportional hazards regression) is method for investigating the effect of several variables upon the time a specified event takes to happen (Kleinbaum 2005). The method does not assume any particular "survival model" but it is not truly nonparametric because it does assume that the effects of the predictor variables upon survival are constant over time and are additive in one scale (Cox 1975).

The Cox regression model is also known as semiparametric method because there is no assumption about the distribution of survival times, but it assumes that the effects of different variables on survival are constant over time (proportionality assumption) and additive over a particular scale (Bradburn et al. 2003).

The Cox Proportional model is the most commonly used multivariable approach for analyzing survival data in medical research. It is essentially a time-to-event regression model, which describes the relation between the event incidence, as expressed by the hazard function, and a set of covariates, it is considered a semi-parametric approach because the model contains a non-parametric component and a parametric component. [Ng'andu. 1997].

Provided that the assumptions of Cox regression are met, this function will provide better estimates of survival probabilities and cumulative hazard than those provided by the Kaplan-Meier function (bradbum 2003).

2.2. Hazard Ratios

Cumulative hazard at a time t is the risk of dying between time 0 and time t, and the survivor function at time t is the probability of surviving to time t.

Often the Cox regression model is called the proportional hazards (PH) model. The reason for this is that the standard model assumes that the excess risk (the vertical distance between the log of the incidence rate of both groups) between the two groups is constant throughout the follow-up time. This fact is the reason that the hazard ratio has such a simple and useful interpretation. Although in practice it is unlikely that the proportional hazards assumption is ever fully satisfied, important violation of the PH assumption may result in wrong and misleading estimates (Paul et al 2008).

The coefficients in a Cox regression relate to hazard; a positive coefficient indicates a worse prognosis and a negative coefficient indicates a protective effect of the variable with which it is associated.

The hazards ratio associated with a predictor variable is given by the exponent of its coefficient; this is given with a confidence interval under the "coefficient details" option in Stats Direct. The hazards ratio may also be thought of as the relative death rate, (Armitage & Berry 1994). The interpretation of the hazards ratio depends upon the measurement scale of the predictor variable in question, (Sahai & Kurshid 1996) for further information on relative risk of hazards.

2.3. What is the time-to-event (TTE) data?

Time-to-event (TTE) data is unique because the outcome of interest is not only whether or not an event occurred, but also when that event occurred. Traditional methods of logistic and linear regression are not suited to be able to include both the event and time aspects as the outcome in the model (Vittinghoff et al 2012). Traditional regression methods also are not equipped to handle censoring, a special type of missing data that occurs in time-to-event analyses when subjects do not experience the event of interest during the follow-up time. In the presence of censoring, the true time to event is underestimated. Special techniques for TTE data (Hosmer et al 2008), as will be discussed below, have been developed to utilize the partial information on each subject with censored data and provide unbiased survival estimates. These techniques incorporate data from multiple time points across subjects and can be used to directly calculate rates, time ratios, and hazard ratios (Kleinbaum & Klein, 2012).

2.4. Model analysis and deviance

A test of the overall statistical significance of the model is given under the "model analysis" option. Here the likelihood chi-square statistic is calculated by comparing the deviance (- 2 * log likelihood) of your model, with all of the covariates you have specified, against the model with all covariates dropped (Rodrfguez 2001). The individual contribution of covariates to the model can be accessed from the significance test given with each coefficient in the main output; this assumes a reasonably large sample size.

Deviance is minus twice the log of the likelihood ratio for models fitted by maximum likelihood (Hosmer and Lemeshow, 1989 and 1999; Cox and Snell, 1989; Pregibon, 1981). The value of adding a parameter to a Cox model is tested by subtracting the deviance of the model with the new parameter from the deviance of the model without the new parameter, the difference is then tested against a chi-square distribution with degrees of freedom equal to the difference between the degrees of freedom of the old and new models. The model analysis option tests the model you specify against a model with only one parameter, the intercept; this tests the combined value of the specified predictors/covariates in the model (Lee & Wang 2003).

2.5. Cox proportional hazards model

All statistical techniques intended for use in the study's practical application are discussed here, with the Cox proportional hazards model, the part involving the determination of the survival function and the risk function, and the comparison of diseases associated with renal failure or more survival data serving as the study's primary focus in this area.

The basic risk function h_0 (t), which is the Cox model's first term and is only connected to time, has also been estimated in this section. This estimation is the first and fundamental step in achieving the Cox model's maximum plausibility.

This function made it clear how the risks of the event change as time passes. In mathematical terms, this explains the behavior of the random time variable (T), which is why it must be a non-negative random variable. As a result, any probabilistic model that is defined on the non-negative field [0,∞) can be representative of the shape of the fundamental risk function. (Al-Tanji 2014).

2.6. The following functions describe the residence period

Suppose (T) represents a non-negative continuous random variable and represents the waiting time (retention time) from the moment of observing the element until the occurrence of the event to be studied (Base& Manning 2004).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 fonction providers the likelihood That an élément Will romain in the community for the specified amont of time (t) or the likelihood That the element Will pass Awa within That time (o,t).

The survival function, as indicated in the formula (2) below, Can alsobé used to represent the residence time (T):

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

And This functiongivestwopossibilities (Zhou 2004) eitherthat the survival time exceeds the time moment (t), or the possibilitythat the absence of danger occurs in the time period (o,t).

The survival function, which is typically derived from the distribution function, can be represented as a density function to explain the survival time:

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

Since the continuous random variable (T) (T) Isdefined over the period (0,∞), then equation (3) can be rewritten in the form:

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

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

The survival function is usually used in survival analysis, although the probability distribution function is more commonly used in other fields of statistics (kleinbaum 2005). It Is noted that the survival function has the following properties :

It is a non-increasing function for each t>0, Thatis:

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

Whereas equation (6) can be rewritten to become:

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

From equation (7) we get:

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

At the beginning of the study, i.e. time zero, all the studied elements are still alive, and this is expressed by the following relationship:

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

At the end of time, the survival function becomes:

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

2.7. Hazard Function

The hazard function h(t) is defined as the conditional probability that the hazard will occur to an element at a moment in time t, knowing that this element is still surviving until the same moment in time (t). It is usually given by formulas (kleinbaum 2012) :

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta T \setminus T > t)}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t, T < t)}{P(T > t).\Delta t} = \frac{f(t)}{s(t)} \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

And since (f(t) is the derivative of the function S(t), the risk function can be rewritten to become:

$$h(t) - \frac{d}{dt} InS(t) \ldots\ldots\ldots\ldots(12)$$

We obtain the following results by integrating both sides of the equation for the interval [0,t]:

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

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

The quantity is known as:

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

The function of rising hazards, which is represented by the equation derived in (15), represents the risks that an element in the examined community has to deal with from time (0) to time (t) (vittinghoff, 2012).

2.8. Expected Lifetime

Suppose (μ) represents the average residence time of an element and gives:

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

Since -f(t) is the derivative of S(t) and based on the equations given in (9) and (10), we can integrate (Bemmaor & Glady 2011) to get the expected time (T).

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

The value of the expectation for any element that has survived until the moment in time (t_0) is given by the formula:

$$E[T - t_0 \setminus T \geq t_0] \ldots\ldots\ldots\ldots(18)$$

Such as:

$$E[T - t_0 \setminus T \geq t_0] = \int_0^\infty t\ dp(T - t_0 \leq t \setminus T > t_0) = \int_0^\infty t\ dp(T \leq t_0 + t \setminus T > t_0) \ldots\ldots\ldots(19)$$

The remaining portion of the distribution function for the survival time is thus provided by:

$$p\left(T \le \mathrm{t}_0 + t \setminus T > \mathrm{t}_0\right) = \frac{p\left(\mathrm{t}_0 < T \le \mathrm{t}_0 + t\right)}{p\left(T > \mathrm{t}_0\right)} = \frac{F\left(\mathrm{t}_0 + t\right) - F + \left(\mathrm{t}_0\right)}{S\left(\mathrm{t}_0\right)} \text{.....(20)}$$

And this is how density is expressed:

$$\frac{d}{dt}\frac{F\left(\mathrm{t}_0 + t\right) - F\left(\mathrm{t}_0\right)}{S\left(\mathrm{t}_0\right)} = f\ \frac{\left(\mathrm{t}_0 + t\right)}{S\left(\mathrm{t}_0\right)} \text{.....(21)}$$

From (Pena 2003) comes the estimated residency time:

$$E\left[T - \mathrm{t}_0 \setminus T \ge \mathrm{t}_0\right] = \frac{1}{S\left(\mathrm{t}_0\right)}\int_0^{\infty} t\ f\ \left(t + \mathrm{t}_0\right)dt\ \text{).....(22)}$$

2.9. Most often used survival models

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

Table 1 Includes the most important and well-known survival distributions

Distributed probability	function of life	mediators
Exponential	$e^{-\lambda t}$	$\lambda > 0$
Natural logarithmic	$1 - \emptyset = \frac{int - u}{\sigma}$	$\mu \in \mathrm{R}, \sigma > 0$
Logistics	$\left[(1 + (t/a)^{\beta}\right] - 1$	$a >, \beta > 0$
Weibull	$e^{-(\lambda_1)\gamma}$	$\lambda > 0, y > 0$
Gompertz	$e^{\theta^{2\left(1 - e^{\theta t}\right)}}$	$\lambda > 0, \theta \in R$

Sources: Machin 2008

The exponential distribution, logistic, computer, and Weibull distribution are considered one of the most important and most famous probability distributions for 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.

2.10. Exponential Distribution

Although it just has one coefficient and is one of the simplest distributions for expressing residence time, the exponential distribution has several uses in all facets of daily life (Balakrishnan 1996). State that the probability density function for the exponential distribution is as follows:

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

Note that ($0 > \lambda$) is the parameter of the distribution. The probability distribution function is given by the formula:

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

And moments in an exponential distribution are given by the formula:

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

By changing the numbers in equation (25) above, which calculates the mean, one can obtain measures of central tendency (Eckford 2016):

And by substituting (r = 1) into equation No. (25), we get the prediction as in equation (26) below:

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

As for finding the median, we solve equation $F(t)=\frac{1}{2}$ to find that the median is equal to $\frac{In(2)}{\lambda}$. We do not find the variance as one of the measures of dispersion; we get it through:

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

As for the skewness, it can be obtained from the formula:

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

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

2.11. Exponential Survival function

The survival function for an exponential distribution is given as (Kleinbaum 2005):

$$S(t)=e^{-\lambda t},t\geq 0\text{.....(29)}$$

That is, $(e^{-\lambda t})$ represents the probability that any element in the community will survive for the entire time period (O,t) at least.

Returning to equation (10), we find:

$$h(t)=\lambda\text{.....(30)}$$

The exponential distribution function is fixed for each (0,t), and thus represents the rate at which the event occurs at each time. Since the rate of hazard varies over time in situations like machine failure, where the likelihood of failure rises with time, it is evident that the exponential distribution does not agree in these situations. However, there are some circumstances where the death rate can be constant. For instance, the death rate (breaking) of glassware is thought to be constant because the rate of death (breaking) of the vessels is fixed, i.e., it is unaffected by the passing of time, which is expressed by the relationship:

$$P(T>s+t\backslash T>s)=P(T>t);\forall s,t\geq 0\text{.....(31)}$$

2.12. Gompertz Model

This model is one of the commonly used models in survival analysis due to the flexibility of the distribution, where the density function can take different forms according to the different values of the mediators.

This model corresponds to many biological phenomena, and it was also found that it corresponds to the retention time of customers (Easton 2009). The probability density function takes the form:

$$f(t) = abe^{bt}e^{a}e^{-ae^{bt}}, t \geq 0, a > 0, b > 0 \ldots\ldots\ldots\ldots(32)$$

The median (a) is called the figure median, and (b) the scale median.

The distribution function has the following form:

$$(T) = 1 - e - aebt, > 0 \ldots\ldots\ldots\ldots(33)$$

3. Results

Hemodialysis is the major treatment modality for renal failure therapy with the highest mortality rates, followed by kidney transplantation and peritoneal dialysis (Carrero et al 2018, Kainz et al 2019). For these reasons, one of the most worrying outcomes of RF is end-stage renal disease, in which there is a need for renal failure therapy. The death of the patient was the event studied. Transplantation and recovery of renal function were censored observations. For all patients, the covariates disease duration, blood type, sex, residence, family history, parents related, marital status, job, age and other related disease were collected at the start of renal replacement therapy.

In this paper, we focused in how to use KM and Cox regression models and visual field progression in patients with renal failure disease (RFD). 130 patients with RFD were followed up for at least 20 years. In these period, 64(49%) approximately half of sample follow-up within 6 – 10 years with suspected (RFD). The most frequently patients had O+ and A+ blood type with 43(33%) and 46(35%) respectively, and the influence of several clinical and demographic variables on patient survival. The main baseline data set collected were: sex, age, race, occupation, marital status, length of treatment, and hemodialysis characteristics.

Table 2 Demographical and clinical characteristics

Variable		**Frequency**	**Percentage**
Blood type	A+	43	33
	B+	23	18
	AB+	6	4.6
	O+	46	35
	A-	6	4.6
	B-	2	1.5
	AB-	2	1.5
	O-	2	1.5
Sex	Male	64	49
	female	66	51
Housing	city	59	45
	village	71	56
family medical history	previous injuries in the family	78	60
	no previous injuries in the family	52	40

Relationship between the parents	Parents are relatives	97	75
	Parents are not related	33	25
Marital status	Unmarried	62	48
	married	47	36
	Other	21	16
Jobs	employee	41	32
	retired	12	9
	free business	9	7
	unemployed	49	38
	House wife	7	5
	Student	12	9
The duration of the disease by years	1-5	41	32
	6-10	64	49
	11-15	24	18
	16-20	1	1
Ages of renal failure	15- 24	12	9
	25- 34	19	15
	35- 44	42	32
	45- 54	29	22
	55- 64	18	14
	65- 74	7	6
	75+	3	2

From the above table, we found that all most the years of kidney failure (6-10) years was 49%, and the more related blood types of people with kidney failure, where (O + : 43%) (A + : 32%) (B + : 18%), and (AB+: 5%) respectively, and 51% of the sample are females, 49% males, with ranged age (35-45), 32%, and 56% of suspected sample of renal failure residence at village while 45% residence town and cites, 60% of sample had do not have a previous injury to the disease, 40% only have a previous injury to the disease, and about is any related between suspected and have parents 75% was relatives and 25% without, and approximately was 48% of the patients are unmarried, 36% married, and 38% are unemployed, 32% are employees.

Table 3 Diseases associated with kidney failure

Statement	**Injured**	**un Injured**	**total**
other diseases	62 48%	68 52%	130
Having diabetes	59 45%	71 55%	130
heart disease	105 81%	25 19%	130
delayed growth	62 48%	68 52%	130

osteoporosis injury	55 42%	75 58%	130
Hepatitis infection	97 75%	33 25%	130

The table shows the diseases associated relate with renal failure, 48% of the sample have other diseases, while 52% are not infected, as well as 55% do not have diabetes not need to injured, just 45% have renal failure patients need injured, for the related renal failure with heart disease 81% have a heart disease need to injured renal failure while only 19% do not, 48% of renal failure patients have growth retardation, 52% do not have, and 42% patients have osteoporosis, while 58% do not have, 75% of the sample are infected with it, while only 25% are not infected with hepatitis C.

3.1. The results of the Kaplan-Meier method

The Kaplan-Meier method was used to analyze the data individually and to calculate the cumulative survival rate and the Log-Rank test, to extract factors that impacted the survival rate and examine the data, and to shows the significance of each of the study variables. In addition, Cox's proportional hazards model was used with heomdialysis treatment as a time-dependent exposure to calculate risk rate.

Table 4 Results of Kaplan-Meier analysis for all variables

Variable		**Event Views**	**Censored Views**	**Chi square**	**P-Value**
Blood type	A+	43	25	11.541	0.117
	B+	23	7		
	AB+	6	4		
	O+	46	13		
	A-	6	1		
	B-	2	0		
	AB-	2	1		
	O-	2	0		
Sex	Male	64	32	0.883	0.364
	female	66	22		
Housing	city	59	15	0.665	0.543
	village	71	39		
family medical history	There are previous injuries in the family	78	26	59.701	0.000
	There are no previous injuries in the family	52	28		
Relationship between the parents	Parents are relatives	97	43	0.856	0.452
	Parents are not related	33	11		
Marital status	Unmarried	62	39	6.634	0.368
	married	47	12		
	Other	21	2		
Jobs	employee	41	19	0.023	0.857
	retired	12	5		

	free business	9	2		
	unemployed	49	22		
	House wife	7	1		
	Student	12	4		
other diseases	Injured	62	35	62.091	0.000
	un Injured	68	19		
Having diabetes	Injured	59	41	74.583	0.000
	un Injured	71	13		
heart disease	Injured	105	50	0.203	0.583
	un Injured	25	38		
delayed growth	Injured	62	15	55.002	0.000
	un Injured	68	30		
osteoporosis injury	Injured	55	40	20.473	0.000
	un Injured	75	31		
Hepatitis infection	Injured	97	55	64.730	0.000
	un Injured	33	12		

Source: prepared by the researcher using the program (stata)

Through the proven results in the previous table, it is clear that the variables (blood type, sex, Housing, Relationship between the parents, marital status, jobs, and heart disease) have no significant effects on the renal failure survival time. Age at discovery of the disease, family medical history, other diseases, having diabetes, delayed growth, osteoporosis injury, Hepatitis infection) have been shown to have significant effects on the patient's survival time.

3.2. Cox regression model results

Where in this paragraph the data is analyzed in two ways, the first is the analysis of all the study variables included in the data presentation, and the second method is the analysis of the study's moral data only as in Table No. (4) the results of Kaplan-Meier analysis.

3.3. The Cox regression method results

In this method, we used Backward Method to estimate the significance variables to build the model, as shown in the following table:

Table 5 Estimating parameters of Cox regression model

variable symbol	**Variables**	β_i	$Exp(\beta_i)$	**standard error**	**Wald statistics**	**d.f**	**P.value**
X3	The duration of the disease	0.003	1.008	0.001	96.365	1	0.000
X4	Housing	0.725	1.923	0.189	10.328	1	0.002
X10	Having diabetes	1.233	4.522	0.342	13.679	1	0.000
X14	Hepatitis infection	0.694	1.849	0.276	4.062	1	0.034

The source was prepared by the researcher using the program (stata)

The mathematical equation of the Cox model can be written with the estimated parameters (β_i) as in the following equation. (34):

$$h(t/x_i) = h_0 \exp\left[0.003x_3 + 0.725x_4 + 1.233x_{10} + 0.694x_{14}\right].....(34)$$

$$sig \quad = \quad 0.000 \quad 0.002 \quad 0.000 \quad 0.034$$

As well as to calculate the coefficient of determination (R^2) for the estimated model as a whole, it can be calculated based on the following formula:

$$R^2 = 1 - \exp\left[\frac{2}{130}\left(-535.056 + 435.621\right)\right] = 0.657.....(35)$$

Whereas, interpreted as 65.7% of the independent variables determine the survival time of the estimated model.

3.4. The results of the analysis of the study's significant variables only, by Cox's method

The researcher has made an estimated model based on the significant variables only in the Kaplan-Meier method in Table (4) which are (7) significant variables and by adopting the Backward Method to estimate the parameters of the model as in the table (6):

Table 6 Results of estimating the parameters of the Cox regression model

variable symbol	Variables	β_i	$Exp(\beta_i)$	standard error	Wald statistics	d.f	P.value
X3	The duration of the disease	0.003	1.008	0.001	96.365	1	0.000
X10	Having diabetes	1.233	4.522	0.342	13.679	1	0.000

The table shows the results obtained, including the parameters of the estimated final model, as it contains the final formula with two significant variables, and the regression equation can be written as follows:

$$h(t/x_i) = h_0 \exp\left[0.003x_3 + 1.233x_{10}\right].....(36)$$

$$sig \quad = \quad 0.000 \quad 0.000$$

The value of the coefficient of determination (R^2) for the estimated model is:

$$R^2 = 1 - \exp\left[\frac{2}{130}(-534.813 + 414.082)\right] = 0.727.....(37)$$

This means that the independent variables contribute to the survival rate of 72.7%.

4. Conclusions

The study concluded that the Kaplan-Meier method for determining the significant variables proved its effectiveness on the estimated model (Cox model), and the estimated model includes all variables whose significance was shown by the Kaplan-Meier method has a clear preference over the proposed model through the coefficient of determination, and (age when the injury was discovered) is of great importance and a major role in the duration of survival of patients with renal failure, also (incidence of diabetes) is less important than the age variable in the duration of survival time. The necessity of using the Kaplan-Meier method in such cases to determine the importance of the independent variables, which can play an important role in denoting the dependent variable before starting the analysis processes related to the survival time. It recommended to backward any variables no significant. Also be careful to review the patients' data and accurately record every small and large number of patients.

Compliance with ethical standards

Acknowledgments

The authors would like to thank the police hospital for providing data.

Disclosure of conflict of interest

No conflict of interest.

References

[1] Hassan, Raad Fadel, and Saleh, Aida Hadi, "Using the Cox regression model for the survival times of patients with brain cancer in Iraq", a published study, Journal of Al-Rafidain University College of Science, Issue 34, pp. 40-61, 2014.

[2] Lee, T. & Wang, W., " Statistical Methods for Survival Data Analysis",2nd ed., Wiley, New York, 2003.

[3] Rodrfguez, G," Generalized linear models " , Princeton university ,USA, 2001 .

[4] Rodrfguez, G., "Survival Models",Chapter 7, Princeton university, USA , 2010. 6-Fox , Y, " cox proportional hazard regression for survival data", 2002.

[5] Kleinbaum K. M, "Survival Analysis A Self-Learning Text", 2Ed. Springer, 2005.

[6] Cox, D. R , "Partial likelihood" , Biometrika , Vol-62 ,No-3, P(269- 276) , 1975.

[7] Hosmer, D. & Lemeshow, S. & May, R.," Applied Survival Analysis: Regression Modeling of Time to Event Data" , 2nd edition Wiley , New York ,USA , 2007.

[8] Collett, D.,"Modeling Survival Data in Medical" , Research ,Chapman and Hall London , 2003.

[9] Mantel, N," Evaluation of survival data and two new rank order statistics arising in its consideration", Cancer Chemotherapy Report, 1966.

[10] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival analysis part II: multivariate data analysis – an introduction to concepts and methods," British Journal of Cancer, vol. 89, no. 3, pp. 431–436, 2003. View at: Publisher Site | Google Scholar

[11] Ng'andu NH, An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model. Stat Med 16: 611–626. 1997, PMID 9131751.

[12] Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics, 2nd New York, 2012, NY: Springer.

[13] Hosmer DW, Lemeshow S, May S. Applied Survival Analysis: Regression Modeling of Time-to-Event Data, 2nd ed. 2008, Hoboken, NJ: John Wiley & Sons, Inc.

[14] Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text, 3rd ed. 2012, New York, NY: Springer Science + Business Media, LLC.

[15] M. K. Goel, P. Khanna, and J. Kishore, "Understanding survival analysis: Kaplan-Meier estimate," International Journal of Ayurveda Research, vol. 1, no. 4, pp. 274–278, 2010. View at: Publisher Site | Google Scholar.

[16] Kitty J. Jager, Paul C. van Dijk, Carmine Zoccali, Friedo W. Dekker, The analysis of survival data: the Kaplan–Meier method, Kidney International, Volume 74, Issue 5, 2008,Pages 560-565,

[17] Carrero JJ, Hecking M, Chesnaye NC, Jager KJ. Sex and gender disparities in the epidemiology and outcomes of chronic kidney disease. Nat Rev Nephrol [Internet]. 2018;14(3):151–64.

[18] Kainz A, Berner C, Ristl R, Simon A, Stamm T, Zitt E, et al. Sex-specific analysis of haemodialysis prevalence, practices and mortality over time: the Austrian Dialysis registry from 1965 to 2014. Nephrol Dial Transplant. 2019; 34(6):1026–35.

[19] Paul C. van Diik, Kitty J. Jager, Aeilko H.Zwinderman, Carmine Zoccali, Friedo W. Dekker. The analysis of survival data in nephrology: basic concepts and methods of Cox regression, ABC epidemiology, September 02, 2008, 74(6): P 705 – 709.

[20] D. Easton. Complementary Gompertz survival models: Decreasing alive versus increasing dead. J. Gerontol. (2009) 64A, 550–555.

[21] Balakrishnan, N. and Basu, A. P. The Exponential Distribution: Theory, Methods, and Applications. New York: Gordon and Breach, 1996.

[22] Eckford, Andrew W.; Thomas, Peter J. "Entropy of the sum of two independent, non-identically-distributed exponential random variables". arXiv:1609.02911, 2016