EJSIT
European Journal of Science, Innovation and Technology
ISSN: 2786-4936
www.ejsit-journal.com
Volume 3 | Number 1 | 2023

Estimation of the Cox Proportional Hazards Model Using the Most Important Variables Affecting the Risk of Neonatal Death in Sudan

Marwa Yousif Ahmed[*], Afra Hashim, Elsamoual Mohammed Kurtokeila
Department of Statistics, College of Science,
Sudan University of Science and Technology, Sudan

Abstract

This paper aimed to identify the important factors that affected neonatal mortality and estimate Cox proportional hazards model with these factors. Data were collected from Omdurman Maternity Hospital, from the record of pregnant women from the first follow-up until the delivery and whether the neonate was alive or dead. The data focused on demographic variables (mother's age, number of previous delivery, city, number of neonate and the sex of neonate) and health variables (neonate weight and mode of delivery). Through log rank test the variables (age, previous delivers and weight of neonate) had a significant effect and the estimated single Cox proportional hazards models with these significant variables are significant and also the multiple Cox proportional hazards model is significant.

Keywords: Explanatory Variables, Neonatal, Hazard Function, Survival Function, Cox Proportion Hazard Model

Introduction

The study of mortality occupies a special place in the field of demographic research, as it represents the negative element of population growth, and its decline is mainly related to the extent of social and economic progress achieved by society, and attention has been directed to the importance of studying mortality, especially neonatal.

Studying the determinants of neonatal mortality, the dependent variable in which is the time that passes until death occurs, so it is considered one of the studies that can apply the cox proportional hazards model to it. This study came to identify the most important factors that can affect neonatal mortality in Sudan, which this factors to include in to cox proportion hazard model and evaluate the model.

Data & Methods

Data

Data were collected from pregnancy follow-up files and the birth records from Omdurman Maternity Hospital in 2020; during this period, delivers will be recorded. The date of the first follow-up of the pregnant mother, and the date of delivery and an indication of whether the newborn (neonate) was alive or dead. Accordingly, the variables were divided into dependent variable is (time to event) the time the mother spends in study until delivery and explanatory variables age (mother's age), previous deliveries (number of previous deliveries to the mother), city (area of residence), number of neonate (twin - single), the sex neonate (male - female), neonate weight (normal (2.5-4 kg), abnormal), mode of delivery (normal vaginal delivery).

[*] Corresponding Author

Methods

The Kaplan-Meier estimate of the survival function

In the past, the statistical methods used to study human death were the method of life tables, but these methods are no longer important after the development of statistical methods such as the Kaplan-Meyer method (Kaplan Meier) and abbreviated as (KM). That is destined (KM) for the survival curve is usually used for the analysis of individual data, as the method of life tables is used for the collected data, and because it is a method for counting the collected data, therefore it is inaccurate like the (KM) method, which uses the values of the items and that this estimator contributes to all the items observed (monitored and unsupervised). (Times of death) by taking survival at any point as a series of steps defining the observed survival times and death times, and using the observed data to estimate the conditional probability of survival at any time of death and multiplying the values of these probabilities to obtain the estimated survival function.

Suppose we have (r) is singular and that:

$$0 \leq t_1 \leq t_2 < \cdots < t_r < \infty$$

are the ordered times of death and that (r_j) is the size of the risk group at (t_j) and (d_j) indicates the number of deaths observed at j =, tj$1.2.\ldots r$

The Kaplan-Meyer estimator of the survival functions (t) gives my clate:

$$\hat{s}(t) = \prod_{j=1}^{k}\left[1 - \frac{d_j}{r_j}\right] \qquad (1)$$

And that this estimator is a scalar function whose values only change with each death time and is sometimes known as the marginal product estimator.

Kaplan-Meier estimate of the hazard function

The risk function is important when defining survival data regression models, and the word risk refers to describing the concept of death in a period after time.t) provided that the item remained alive until time (t).

The Kaplan-Meier estimator of the risk function is found by taking the ratio of the number of deaths at a specific time of death to the number of items at risk at that time. Assuming that the risk function will be constant between the two successive times of death, the risk per time unit can be found by dividing by the time period, and therefore if We assumed that it refers to the number of deaths at the time of death (d_iJ) and that $t_i = 1.2.\ldots.r$, and if (r_j) refer to the number of items in risk at time (t_j), then the risk function for the period from (t_j) to (t_{j+1}) can be estimated through the formula:

$$\hat{h}(t) = \frac{d_j}{t_j \tau_j} \qquad (2)$$

$$\text{for } t_j \leq t < t_{j+1}$$

$$\tau_j = t_{j-1} - t_j$$

And that the approximate standard error of the function$\hat{h}_{(t)}$ can be found from the varianced_i, and assuming that it follows the binomial distribution with the parameters (p_i, r_i), so that it is (p_i) the probability of death in the period of length (T) so:

$$\operatorname{var}(d_j) r_j p_j (1 - p_j)$$

With an estimate that: $P_i \left[\frac{d_i}{r_i}\right]$

$$\operatorname{Se}\left(\hat{h}(t)\right) = \hat{h}(t)\sqrt{\frac{r_j - d_j}{r_j d_j}} \qquad (3)$$

[Collett, 2003, p. 31]

Semi-parametric Cox regression model for survival data analysis

When analyzing survival data, the concern is about the risk of death at any time after the original time of the study. As a result, the risk function is modeled directly in the survival analysis, and there are two reasons for modeling survival data. Risk, the second reason to model the risk function is to get the risk function itself for the individual (Collett, 2003).

The basic model that we will discuss is the Cox model of relative risk or the risk model. Cox proposed this model (1972). This model is based on the assumptions of relative risk. This model does not assume a specific form for the probability distribution of survival times. Therefore, this model is known as a semi-parametric model and gives the following:

$$h(t, x) = h_0(t)\, e^{Bx} \tag{4}$$

e^{BX} – it is a function of the values of the variables X and B are model parameters, which is the parametric part in the model and the ratio of hazards function for a single variable (x) that has values and is usually used to compare two groups:$[x_0.\, x_1]$

$$HR(t.\, x) = \frac{h(t.\, x_1)}{h(t.\, x_0)}$$

$$= \frac{h_0(t)e^{Bx_0}}{h_0(t)e^{Bx_0}} = \frac{e^{Bx_0}}{e^{Bx_0}} = e^{B(x_1 - x_0)} \tag{5}$$

We note that the risk level does not depend on time, and therefore the model is known as the relative risk model because this level is fixed with time.

Since the survival function can be found from the risk function, and then the survival function is:

$$s(t.\, x) = e^{-H(t.x)} \tag{6}$$

Results

The following tables and figures are the results of analysis the data.

Table 1: Log rank test for explanatory variables

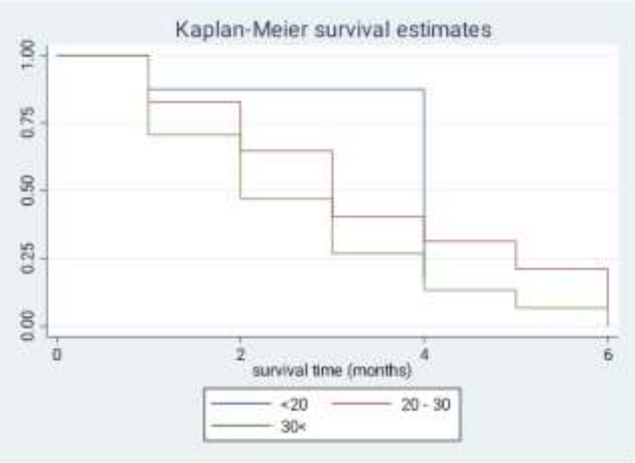
Log rank test	Age	Previous delivery	Sex neonate	Number of neonate	City	Weight neonate	Mode of delivery
Chi^2	2.8	2.46	0.19	0.97	0.75	18.13	0.58
p-value	0.0	0.0	0.66	0.32	0.39	0.0	0.45

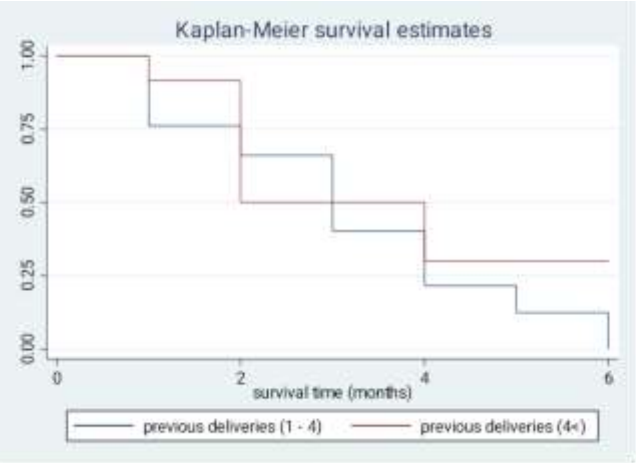
Source: prepared by the researchers by using STATA 17, 2023

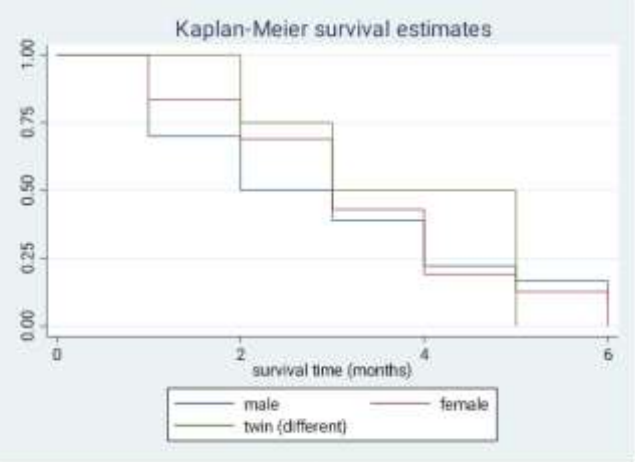
Table 2: Estimated Cox proportional hazards model for each variable

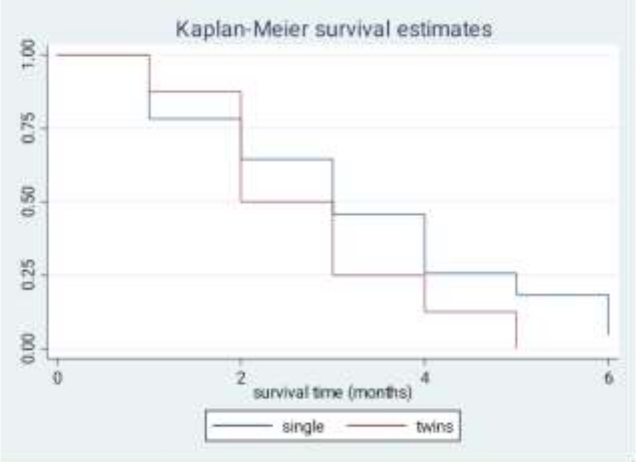
Models	Age	Previous delivery	Sex neonate	Number of neonate	City	Weight neonate	Mode of delivery
Parameter	1.59	1.01	140.71	1.98	1.03	30.77	4.04
Wald test	1.97	1.66	0.44	1.03	0.82	3.98	1.02
p-value	0	0	0.66	0.305	0.44	0	0.31
E .r	0.35	0.30	0.22	0.59	0.34	0.10	0.7

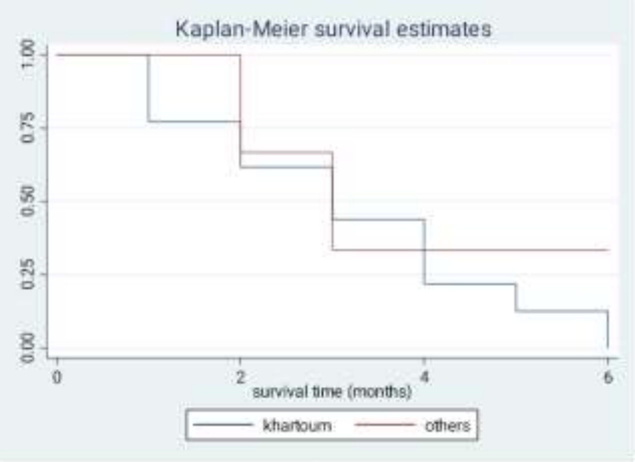
Source: Prepared by the researchers by using STATA 17, 2023

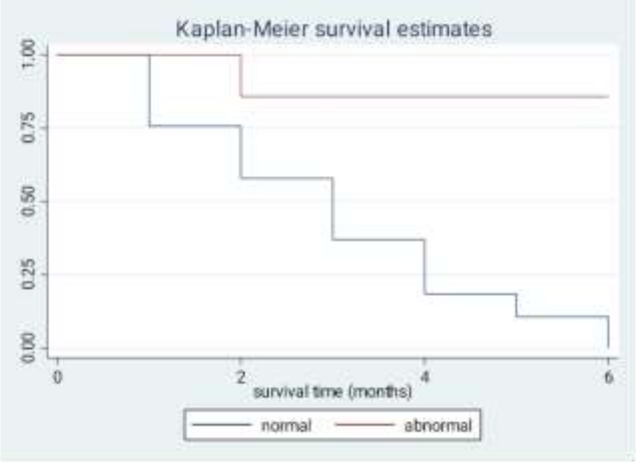

Kaplan-Meier survival estimates
0.00
0.25
0.50
0.75
1.00
0
2
4
6
survival time (months)
<20
20 - 30
30<


Kaplan-Meier survival estimates
0.00
0.25
0.50
0.75
1.00
0
2
4
6
survival time (months)
previous deliveries (1 - 4)
previous deliveries (4<)


Kaplan-Meier survival estimates
0.00
0.25
0.50
0.75
1.00
0
2
4
6
survival time (months)
male
female
twin (different)


Kaplan-Meier survival estimates
0.00
0.25
0.50
0.75
1.00
0
2
4
6
survival time (months)
single
twins


Kaplan-Meier survival estimates
0.00
0.25
0.50
0.75
1.00
0
2
4
6
survival time (months)
khartoum
others


Kaplan-Meier survival estimates
0.00
0.25
0.50
0.75
1.00
0
2
4
6
survival time (months)
normal
abnormal

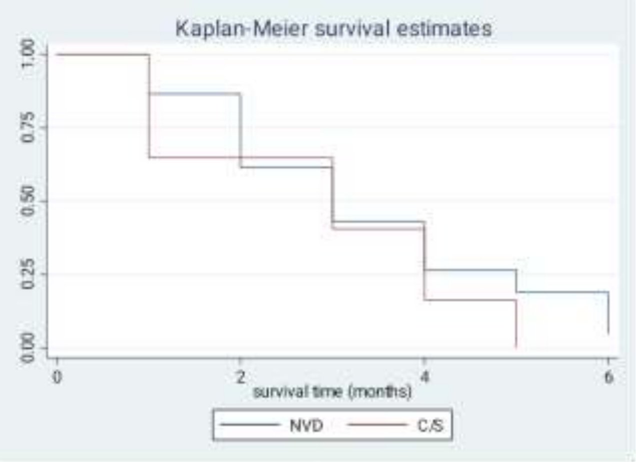


Figure 1: Estimated survival functions for explanatory variables
Source: Prepared by the researchers by using STATA 17, 2023

Table 3: Test significance Cox's multiple model

-2 (log likelihood)	Chi-square test	Probability value
136.37605	8.87	0.00311

Source: Prepared by the researchers by using STATA 17, 2023

Table 4: Estimated coefficients of the Cox multiple proportional hazards model for neonatal mortality

Variables	**Coefficient**	**Wald test**	**Probability value**	**Hazard ratio**	**Standard error**	**Confidence interval**	
						Upper	**Lower**
Mother's age	0.285	3.6	0.001	0.752	0.309	0.688	-1.257
Previous deliveries	1.29	3.92	0.0	0.275	0.479	2.908	0.328
Weight neonate	3.347	3.98	0.0	28.409	0.143	5.016	1.677

Source: Prepared by the researchers by using STATA 17, 2023

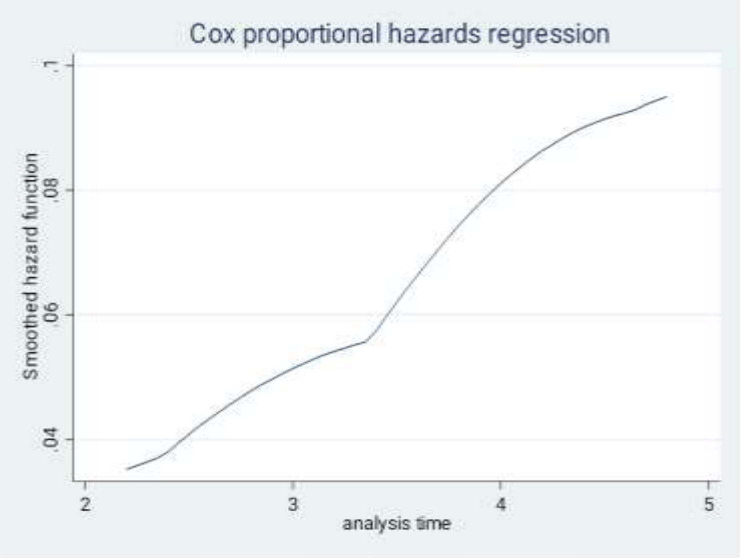


Figure 2: Estimated smoothed hazard function for Cox proportional hazards model
Source: Prepared by the researcher by using STATA 17, 2023

Estimated smoothed hazard function with time can be defined through Table 4.

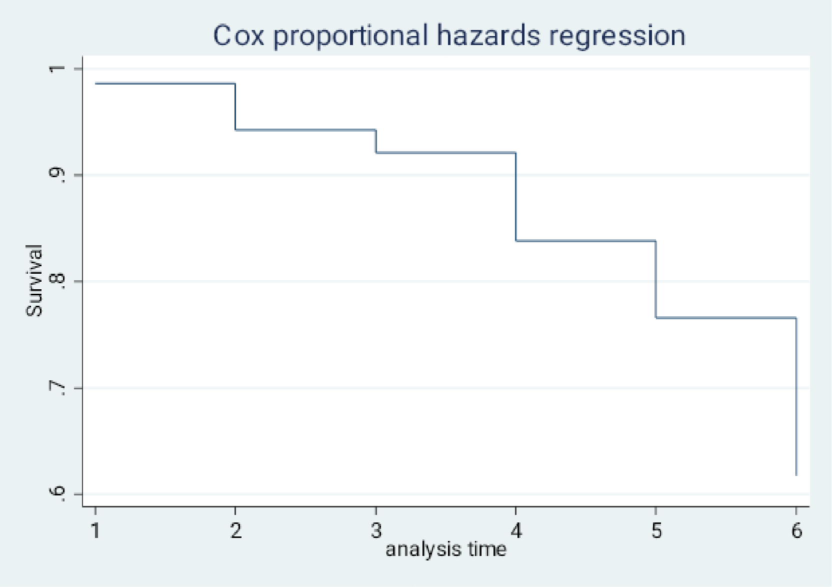


Figure 3: Estimated survival function for Cox proportional hazards model
Source: Prepared by the researcher by using STATA 17, 2023

Discussion

From Table 1 we found that the factors of age (Chi^2 = 2.8, p-value = 0.0), previous delivery (Chi^2 = 2.46, p-value = 0.0) and weight neonate (Chi^2 = 18.13, p-value = 0.0) had a significant effect on the hazard of neonatal mortality in Sudan and the other factors sex neonate (Chi^2 = 0.19, p-value = 0.66), number of neonate (Chi^2 = 0.97, p-value = 0.32), city (Chi^2 = 0.75, p-value = 0.39), mode of delivery (Chi^2 = 0.58, p-value = 0.45) had no significant effect on neonatal mortality in Sudan.

From Table 2 estimation cox proportional hazards models for each variable, model of age (Parameter = 1.59, Wald test = 1.97, p-value = 0) is significant, model of previous delivery (Parameter = 1.01, Wald test = 1.66, p-value = 0) is significant, and model of weight neonate (Parameter = 30.77, Wald test = 3.9, p-value = 0) is significant. In other models there is no significant effect on neonatal mortality in Sudan. Figure 1 showed that the estimated survival function of explanatory variables.

Through Table 1 and Table 2 the variables effecting neonatal mortality in Sudan (Age, previous delivery, Weight neonate) are entered into a semi-parametric proportion hazard model, and the test of significant model.

From Table 3 we found that the value of log likelihood for the model as a whole is (136.37605) and that the value of the chi-square test is (8.87) at the degree of freedom (3) and the probabilistic value (0.00311), this meaning that the model is significant, which indicates that the model can be used in estimating the hazard neonatal mortality in Sudan.

Table 4 showed the results of semi-parametric proportional hazard model estimated for neonatal mortality. The estimated coefficient for the age of the mother is (0.285) and ($p=0.001<0.05$), that means there are significant differences between mother's ages in terms of the risk of neonatal death with the stability of (number of previous delivery and weight neonate), when age changes by one year the hazard ratio decreases by (0.752). The estimated coefficient of the number of previous delivery is (1.29) and ($p=0.000<0.05$), that means there are significant differences in the number of previous delivery in terms of the risk of neonatal death with the stability of other variables (mother's ages and weight neonate), when number of previous delivery changes by one birth the hazard ratio decreases by (0.275). The estimated coefficient of the weight neonate is (3.347) and (p=0.000<0.05) that means the neonate weight has a significant effect on the risk of neonatal death with the stability of mother's ages and number of previous delivery. When weight changes by 1 kg the hazard ratio decreases by (0.752).

References

Collett, D. (2013). *Modelling survival data in medical research*. London: Chapman and Hall.

Cox, D. & Oakes, D. (1984). *Analysis of survival data*. London: Chapman and Hall.

Fleming, T. R. & Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

Guo, S. (2010). *Survival analysis*. New York: Oxford University Press.

Hosmer, D. W. & Lemeshow, S. (1999). *Applied survival analysis: Regression modeling of time-to-event data* (1st ed.). New York: Wiley.

Hougaard, P. (2000). *Analysis of multivariate survival data*. New York: Springer.

Klein, J. P. & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*. New York: Springer.

Lee, E. T., & Wang, J. (2003). *Statistical methods for survival data analysis* (3rd ed.). New York: Wiley.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, *50*(3), 163-170.

McGilchrist, C. A. & Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, *47*(2), 461-466.

Royston, P. & Lambert, P. C. (2011). *Flexible parametric survival analysis using Stata: beyond the Cox model*. College Station, TX: Stata press.

Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, *21*(15), 2175-2197.