Scientific Research Publishing

Open Journal of Statistics, 2022, 12, 532-547
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Log-Rank Test for Comparing Survival Curves of Neonatal Mortality Characteristic Groups in River Nile State-Sudan

Abdellateef Khalifa Hamid Ali[1*], Amin Ibrahim Adam Mohammed[2]

[1]Department of Statistics, Faculty of Mathematical Science, Al-Fashir University, Al Fashir, Sudan
[2]Department of Statistics, Faculty of Economic and Social Studies, Omdurman Islamic University, Omdurman, Sudan
Email: *abdu-53@hotmail.com

How to cite this paper: Ali, A.Kh.H. and Mohammed, A.I.A. (2022) Log-Rank Test for Comparing Survival Curves of Neonatal Mortality Characteristic Groups in River Nile State-Sudan. *Open Journal of Statistics*, **12**, 532-547.
https://doi.org/10.4236/ojs.2022.124032

Received: June 10, 2022
Accepted: August 27, 2022
Published: August 30, 2022

Abstract

This paper concerns the Log-rank test for comparing survival curves of neonatal mortality characteristic groups in River Nile State, Sudan. In this paper, log-rank test is used to compare two or more survival curves for the characteristics of newborn associated with newborn death after using Kaplan-Meier methods to estimate and graph survival curves for the variable of interest as (sex of newborn, weight of newborn, gestational age, mode of delivery and resident type), at the hospital of River Nile state—Sudan, with a sample size 700 of newborn in which the admission to the Neonatal Intensive Care Unit (NICU) of those hospitals during the period 2018-2020. In term of risk of death for newborn we found that 25% of sample study for newborns who were born in River Nile State-Sudan died. In addition, we conclude that after the log-rank statistics and Kaplan-Meier methods were applied, gender does not affect the newborn's risk of survival, while the risk of survival increases when the birth weight is greater than 4.35 kg and the gestational age is greater than 42 weeks. There is no difference in the probability of survival for newborns whether the delivery is normal or cesarean. However, newborns are significantly more likely to survive in urban areas than in rural areas.

Keywords

Neonatal Mortality, Log-Rank Test, Survival Function, Kaplan-Meier Method

1. Introduction

Neonatal mortality is defined as the probability of a baby dying between the first day of birth and 28 days of life. Put differently, all deaths taking place between

the 28 days of life are considered as neonatal mortality [1].

The neonatal period (from birth to 28th day of life) is normally considered to be the most vulnerable and high-risk time in the neonate life because of the highest mortality and morbidity incidence in human life during that period. During this period the neonate risk of death is almost 15 times more than any other time before the first birthday [2].

Approximately 3% to 4% of newborns are born with a major birth defect and will require genetic evaluation.

These birth defects or malformations can be sporadic or associated with other anomalies. Some children may have physical features consistent with a well-known syndrome, while others may have anomalies detected prenatally or postnatal. Other neonatal presentations include some inborn errors of metabolism (acidosis), unexplained seizures, extreme hypotonia, or feeding difficulties. Infants with ambiguous genitalia require a multidisciplinary evaluation involving clinicians from genetics, endocrinology, urology, pediatrics or neonatology, and psychology. A thorough clinical evaluation requires a detailed prenatal history, a family history, and a comprehensive clinical exam, often including anthropometric measurements [3].

Neonatal intensive care unit (NICU) graduates are at high risk for developing hearing loss. When undetected, hearing loss can result in delays in language, communication, and cognitive development [3].

Most neonatal deaths arise in low-income and middle-income countries, and approximately half occur at home. In the past two to three decades, neonatal mortality rates have shown a slow decline in developing countries whereas infant and under-5 mortality rates have declined significantly [2].

Children face the highest risk of dying in their first month of life at an average global rate of 17 deaths per 1000 live births in 2020, down by 54 percent from 37 deaths per 1000 in 1990. In comparison the probability of dying after the first month and before reaching age one day was estimated at 11 deaths per 1000 and the probability of dying after reaching one day and before reaching age 5 weeks was estimated at 9 deaths per 1000 in 2020. Globally, 2.4 million children died in the first month of life in 2020 approximately 6500 neonatal deaths every day with about a third of all neonatal deaths occurring within the first day after birth, and close to three quarters occurring within the first week of life. In 2020, neonatal mortality rate for Sudan was 26.8 deaths per 1,000 live births [1].

Many factors have been linked to the high neonatal mortality rate and these include complications of preterm birth, birth asphyxia, sepsis and pneumonia. The causes of neonatal deaths normally differ from region to region due to the differences in availability of medical resources, cultural practices, and other social issues.

Many studies have shown that neonatal mortality is influenced by multiple factors. Socioeconomics, demographics, the health care system, cultural practices and technology are also important indirect determinants of neonatal mortality.

Considerable researchers are focusing their studies on the neonatal mortality; they deem the count and percentage of the newborn's death at NICU according to several factors affecting newborn live. They found that the residence, mode of delivery and newborn weight are the most factors that cause death event among newborns.

Thus, increasing of neonatal mortality will make the country classified as a developing country.

Improving the health sector for the maternal and fetus, and providing the health services at the rural and urban areas, are important to helping for reducing the event of death of neonates which can't complete their first month.

Without full disclosure of the causes of neonatal deaths, it would be very difficult for government and health authorities to plan and provide appropriate interventions to reduce neonatal deaths in River Nile State. The study therefore examines the factors responsible for neonatal mortality in River Nile State.

The high rates of neonatal mortality among societies are one of the most dangerous factors that cause their deterioration and instability, and the neonatal mortality rate can be used as a measure to judge a country that is considered one of the developed or poor developing countries.

The main purpose of this study is to use Log-rank test to compare several survival curve groups of neonatal mortality data for identifying which groups are more likely to survive, and which factor had high effect on newborns death.

The objective of this paper is to afford the factors which increase the rate of neonatal mortality, and which variable of interest groups had a high probability to survive, through estimating survival curve for the groups of the variables of interest.

The organizational structure of this paper shows as: Introduction, Materials & Methods, Results & Discussion, Conclusion and Recommendations.

2. Materials and Methods

2.1. Materials

The data for this study was extracted from NICU of hospitals in River Nile State about neonate's data from 2018 to 2020. However, it suffices to declare that the NICU of hospitals in River Nile based on representative sample of newborn sex, mode of delivery and residence who were selected using random sampling technique with sample size 700 newborns. The outcome is the risk of neonatal death. Neonatal mortality is defined as the death of a child before completing one month of age. Therefore, time to death was measured in days and infants who lived beyond 28 days were censored at that time for the purpose of survival analysis.

2.2. Methods

The methodology was followed to suit the objectives of this study as, we carried out a preliminary analysis on the comparison of survival experiences with re-

spect to some variable of interest about newborns characteristics by estimating survival curves using Kaplan-Meier Method and performing log-rank tests, then conduct Log-rank test, can be divided into the steps as follows:

1) Collection.

2) Estimation.

3) Testing: the test conducted in this study are:

- Test if there is a difference between survival curves of newborn data.
- Test if there is a difference between each survival curves for the variables of interest groups.

4) Comparison.

In most bio-statistical contexts involving hypothesis testing, one is interested in assessing whether there is evidence that a particular factor has an effect on some health outcome. In practice, the term "effect" is often used loosely to encompass both situations that truly are the assessment of the effect of a factor, such as a randomized trial of an intervention, and situations involving only the assessment of associations, such as an observational study of a particular exposure [4].

In reliability and survival analysis, the survival or longevity depends on individual characteristics of units/subjects. In general, these characteristics are expressed as a set of explanatory variables (also called stresses or covariates), which are possibly time-dependent [5].

Generally, survival analysis is a collection of statistical procedures for the analysis of data in which the outcome variable of interest is time until an event occurs. By event, we mean death, disease incidence, relapse from remission, or any designated experience of interest that may happen to an individual. When doing a survival analysis, we usually refer to the time variable as survival time. We also typically refer to the event as a failure [6].

Most survival analyses must consider a very important analytical problem called censoring. It is caused by not observing some subjects for the full time till failure (or event). We will consider the patient use case to understand the problem easily. We have records for patients dying from heart attack, but in some situations, it may not be possible to mark the exact time of death. Patients might have died either before or after the marked time value. In this case, it is said that the data is censored. In medical diagnosis, a study involves regular follow-ups with the doctor by the patient. The doctor starts taking notes about the patient's health condition in each follow-up and schedules the next date. Problem occurs when the patient dies in between, after the end of or before the study, and hence censoring occurs. There can be three primary reasons for this:

1) Patient does not have the event (death) before the study ends.

2) Patient left follow-up during the study period.

3) Patient died in the study period.

In all above cases, true survival time is not equal to the observed survival time, as the actual time could not be marked. Depending on three situations, there can

be corresponding three types of censoring [7].

Right Censoring:

Censored data are most common in the time-to-event context. Right-censored data are particularly relevant in that context; we may follow some patients after initiation of a new treatment to see how long they live, but since the trial itself usually is restricted in time, we may not be able to follow them all to their death. Instead, we stop the trial at some specific point in time, and for those still alive at that point, we only know that the survival time is longer than the time we have observed them, not the exact value. There are a number of other reasons why follow-up of a subject may cease before the event of interest has occurred.

For a proper analysis to be conducted, the reason why the event has not been observed in an individual must be independent of his underlying risk for the event (non-informative censoring), or there will be some issues around the interpretation of the results. There will be bias in the results if there is a systematic withdrawal of either high- or low-risk patients [8].

Left Censoring:

Left censoring happens when event has already occurred before the start of the study. For this case, true survival time is less than or equal to the observed time (study end). It is useful in modeling the use cases where patients are doing regular follow ups and the exposure of the disease is considered as an event.

Interval Censoring:

Interval censoring happens when event occurs within the study period in between two possible time limits, and as usual actual time could not be noted. The event can be occurred somewhere between time t_i and t_{i+1}. This scheme is applicable to both death and exposure to disease events. Interval censoring is a combination of both left and right censoring with one limit as infinite.

Sample Dataset Structure:

Any survival analysis-related dataset follows a typical structure. From our discussion so far, we can easily understand that time information should be there at first. Second, event status should be there. It is also known as censoring status. It should be either a Boolean or a status indicator (0 or 1). Value 1 indicates occurrence of the event and it is not censored, where as a 0 value says non-occurrence and it is censored. Survival analysis dataset often contains information about the subject under study [7].

Distribution of the survival time from enrollment or starting point to the event of interest, considered as a random variable, is characterized by either one of two equivalent functions: the survival function and the hazard function. The survival function, denoted by $S(t)$, is defined as the probability that a person or item survives longer than t units of time.

$$S(t) = Pr(T > t) \tag{1}$$

$S(t)$ is also known as the survival rate; for example, if times are in days, $S(2)$ is the two-days survival rate, $S(5)$ is the five-days survival rate, and so on. A graph of $S(t)$ versus t is called a survival curve [9]. Alternatively, it can be said that $S(t)$

gives us the probability of a subject surviving after time t. $S(t)$ is nothing but a probability distribution over time. Theoretically, t ranges from 0 to infinity, and of course, $S(t)$ will have values from 0 to 1. Ideally, survival function is represented by a decreasing smooth curve which begins at $S(t) = 1$ at $t = 0$. In practice, curves generated from real datasets look more broken and stepwise. One important property of survival function is that it is monotonically decreasing, *i.e.*, $S(t_1) < S(t_2)$ where $t_1 > t_2$.

On a practical note, survival chances from an event tend to decrease over time [7]. Since a unit either fails, or survives, and one of these two mutually exclusive alternatives must occur, we have:

$$S(t) = 1 - F(t), \quad F(t) = 1 - S(t) \tag{2}$$

where $F(t)$ is the cumulative distribution function (CDF). If T is a continuous random variable, then $S(t)$ is a continuous, strictly decreasing function.

The survival function is the integral of the probability density function (pdf), $f(t)$, that is:

$$S(t) = \int_t^{\infty} f(x)\mathrm{d}x \tag{3}$$

Thus,

$$f(t) = -\frac{\mathrm{d}S(t)}{\mathrm{d}t}. \quad [10]$$

After providing a description of the overall survival experience in the study, we turn our attention to a comparison of the survival experience in key subgroups. These groups are typically defined by covariates thought to be related to survival such as treatment arms in a clinical trial or other key factors. When comparing groups of subjects, we should begin with a graphical display of the data in each group. In studies of survival time, we should graph the Kaplan-Meier estimator of the survival function for each of the groups [11].

To estimate the survival probability when the incidence rate varies over time, we could use a more complex parametric survival model than the exponential model for a good description of other parametric survival model. However, a more common approach is to use a nonparametric method referred to as the product-limit or Kaplan-Meier estimator [12].

The Kaplan-Meier estimator of the survival function [Kaplan and Meier (1958)], also called the product limit estimator, and is the default estimator used by most software packages. This estimator incorporates information from all the observations available, both uncensored (event times) and censored, by considering survival to any point in time as a series of steps defined at the observed survival and censored times. We use the observed data to estimate the conditional probability of confirmed survival at each observed survival time and then multiply them to obtain an estimate of the overall survival function [11].

The general formula for a KM survival probability at failure time $t_{(j)}$ is:

$$\hat{S}_{\left(t_{(j)}\right)} = \hat{S}_{\left(t_{(j-1)}\right)} \times \hat{P}r\left(T > t_{(j)} / T \geq t_{(j)}\right) \tag{4}$$

This formula gives the probability of surviving past the previous failure time $t_{(j-1)}$, multiplied by the conditional probability of surviving past time $t_{(j)}$, given survival to at least time $t_{(j)}$. The above KM formula can also be expressed as a product limit if we substitute for the survival probability $\hat{S}_{\left(t_{(j-1)}\right)}$, the product of all fractions that estimate the conditional probabilities for failure times $t_{(j-1)}$ and earlier.

$$\hat{S}_{\left(t_{(j-1)}\right)} = \prod_{i=1}^{j-1} \hat{P}r\left(T > t_{(i)} / T \geq t_{(i)}\right) \tag{5}$$

The general expression for the product limit formula for the KM survival estimate is shown together with the general KM formula given earlier. Both expressions are equivalent. as follows:

$$\hat{S}_{\left(t_{(j)}\right)} = \prod_{i=1}^{j} \hat{P}r\left(T > t_{(i)} / T \geq t_{(i)}\right) = \hat{S}_{\left(t_{(j-1)}\right)} \times \hat{P}r\left(T > t_{(i)} / T \geq t_{(i)}\right). \tag{6}$$

[6].

The simplest way of comparing the survival times obtained from two groups of individuals is to plot the corresponding estimates of the two survivor functions on the same axes. Both the log-rank and the Wilcoxon tests can be extended to enable three or more groups of survival data to be compared [13].

The method of comparing groups log rank test is widely used to compare two independent survival functions in the presence of censored observations. Several statisticians contributed to the development of the test procedure (Mantel, Cox, Peto and Peto), and thus there are different names associated [14].

The log rank test, a nonparametric procedure for comparing two or more survival functions, is a test of the null hypothesis that all the survival functions are the same, versus the alternative that at least one survival function differs from the rest. A simple Chi-square test that is very similar to the log rank test. For the Chi-square test, we simply let O_1 be the observed number of deaths in group 1, O_2 the observed number in group 2, O_3 the observed number in group 3, and so on until all the groups have been enumerated, then a Chi-square statistic is determined by computing the expected numbers E_1, E_2, E_3, etc., of deaths in each group. For this calculation to hold, all the groups need to come from the same population of survival times. Then, similar to other, the statistic

$$\chi^2 = \left(O_1 - E_1\right)^2 / E_1 + \left(O_2 - E_2\right)^2 / E_2 + \cdots + \left(O_k - E_k\right)^2 / E_k \tag{7}$$

has approximately a Chi-square distribution with $k-1$ degrees of freedom when the null hypothesis is true [15].

Log-rank test for two groups:

We now describe how to evaluate whether or not KM curves for two or more groups are statistically equivalent. In this section we consider two groups only. The most popular testing method is called the log-rank test. When we state that

two KM curves are statistically equivalent, we mean that, based on a testing procedure that compares the two curves in some overall sense, we do not have evidence to indicate that the true (population) survival curves are different [6].

The log-rank test is a large-sample Chi-square test that uses as its test criterion a statistic that provides an overall comparison of the KM curves being compared. This (log-rank) statistic, like many other statistics used in other kinds of Chi-square tests, makes use of observed versus expected cell counts over categories of outcomes. The categories for the log-rank statistic are defined by each of the ordered failure times for the entire set of data being analyzed [6].

Here, for each ordered failure time, $t_{(j)}$, in the entire set of data, we show the numbers of subjects (m_{ij}) failing at that time, separately by group (i), followed by the numbers of subjects (n_{ij}) in the risk set at that time, also separately by group [6].

We now expand the previous table to include expected cell counts and observed minus expected values for each group at each ordered failure time. The formula for the expected cell counts is shown here for each group. For group 1, this formula computes the expected number at time j (*i.e.*, e_{1j}) as the proportion of the total subjects in both groups who are at risk at time j, that is, $n_{1j}/\left(n_{1j}+n_{2j}\right)$, multiplied by the total number of failures at that time for both groups (*i.e.*, $m_{1j}+m_{2j}$). For group 2, e_{2j} is computed similarly [6].

$$e_{1j}=\frac{n_{1j}}{n_{1j}+n_{2j}}\times m_{1j}+m_{2j}, e_{2j}=\frac{n_{2j}}{n_{1j}+n_{2j}}\times m_{1j}+m_{2j} \tag{8}$$

When two groups are being compared, the log-rank test statistic is formed using the sum of the observed minus expected counts over all failure times for one of the two groups.

$$O_i-E_j=\sum_{j=1}^{k}\left(m_{ij}-e_{ij}\right), i=1,2,k \text{ number of failures} \tag{9}$$

For the two-group case, the log-rank statistic, shown as follows:

$$\text{Log-rank statistics}=\frac{\sum\left(O_i-E_i\right)^2}{Var\left(O_i-E_i\right)} \tag{10}$$

The expression for the estimated variance is shown as follows:

$$Var\left(O_i-E_i\right)=\frac{n_{1j}n_{2j}\left(m_{1j}+m_{1j}\right)\left(n_{1j}+n_{2j}-m_{1j}-m_{1j}\right)}{\left(n_{1j}+n_{2j}\right)^2\left(n_{1j}+n_{2j}-1\right)}, i=1,2 \tag{11}$$

For two groups, the variance formula is the same for each group. This variance formula involves the number in the risk set in each group (n_{ij}) and the number of failures in each group (m_{ij}) at time j. The summation is over all distinct failure times [6].

The null hypothesis being tested is that there is no overall difference between the two survival curves. Under this null hypothesis, the log-rank statistic is approximately Chi-square with one degree of freedom.

Thus, a P-value for the log-rank test is determined from tables of the Chi-square distribution [6].

Log-rank test for several groups:

If the number of groups being three or more ($G > 2$), then the log-rank statistic has approximately a large sample Chi-square distribution with $G-1$ degrees of freedom and can be written as follows:

$$\text{Log-rank statistics} \approx \chi^2 = \frac{\sum\left(O_i - E_i\right)^2}{E_i} \tag{12}$$

The approximate formula previously described involving only observed and expected values without variance or covariance calculations can also be used when there are more than two groups being compared. The observed covariates in this study are classified into several newborns' characteristics, such as the resident type, sex of the child, mode of delivery, birth type. These variables have been selected as suggested by the medical and epidemiological literature to have a significant contribution to the survival or death of neonates [6].

3. Results and Discussion

The Kaplan-Meier curve was plotted to describe the survival probabilities of the neonates and to compare the survival experiences for some selected variables of interest about characteristics of the newborns listed as (sex of newborn sex, newborn weight, gestational age, mode of delivery and resident type). The log-rank test was used to test the hypothesis of no difference between groups of each selected variables characteristics of the newborns.

We can show the frequency distribution of some selected factors with respect to some factors affecting survival of neonates in hospitals of River Nile State-Sudan during the period 2018-2020.

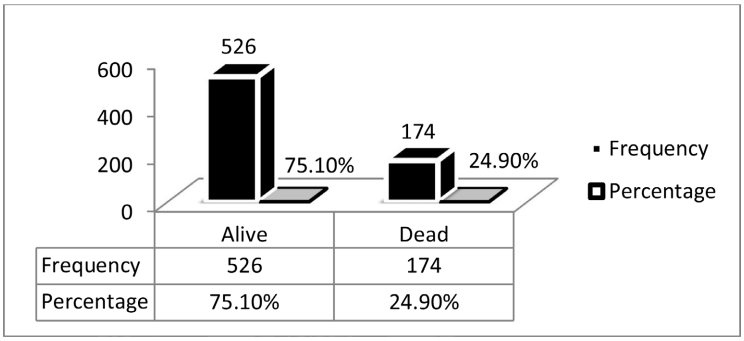
From **Table 1** and **Figure 1**, we conclude that, 75.10% of newborns are alive with frequency 526 and with approximation 24.90% of them are dead with frequency 174. So, we can say the percentage of newborn death is high in study area.

Table 2 shows a summary of the covariates used for the study based on the log-rank test. In term of neonate's characteristics, socio-economic and health care factors, the results show a significant difference in the probability of survival for neonates with respect to sex of newborn, weight of newborn, gestational age, mode of delivery, resident type.

Table 1. Distribution of the study sample according to the Follow up Status.

Follow-up Status	Frequency	Percentage
Alive	526	75%
Dead	174	25%
Total	700	100%

Source: Prepared by the researcher using SPSS.



Source: Prepared by the researcher using SPSS.

Figure 1. Distribution of the study sample according to the Follow up Status.

Table 2. Frequency Distribution of selected factors of interest related to Characteristics of newborns with respect to Survival Status.

Variables of interest	Status				Total	
	Alive		Dead			
	Count	Percentage	Count	Percentage	Count	Percentage
Sex of newborn						
Male	323	73.2%	118	26.8%	441	63.0%
Female	203	78.4%	56	21.6%	259	37.0%
Weight of newborn						
0.35 to less than 2.35	190	63.5%	109	36.5%	299	42.7%
2.35 to less than 4.35	327	83.6%	64	16.4%	391	55.9%
More than 4.35	9	90.0%	1	10.0%	10	1.4%
Gestational Age						
22 to less than 32	52	50.0%	52	50.0%	104	14.9%
32 to less than 42	469	79.4%	122	20.6%	591	84.4%
More than 42	5	100.0%	0	0.0%	5	0.70%
Mode of Delivery						
Normal	271	74.0%	95	26.0%	366	52.3%
Caesarian Section	255	76.3%	79	23.7%	334	47.7%
Resident Type						
Urban	290	80.3%	71	19.7%	361	51.6%
Rural	236	69.6%	103	30.4%	339	48.4%

From **Figure 2** we notice that the KM curve for group 2 is consistently higher than the KM curve for group 1. These figures indicate that group 2, which is the female of newborn group, has better survival prognosis than group 1, the male of newborn group.

From **Figure 3** we notice that the KM curve for group 3 is consistently higher

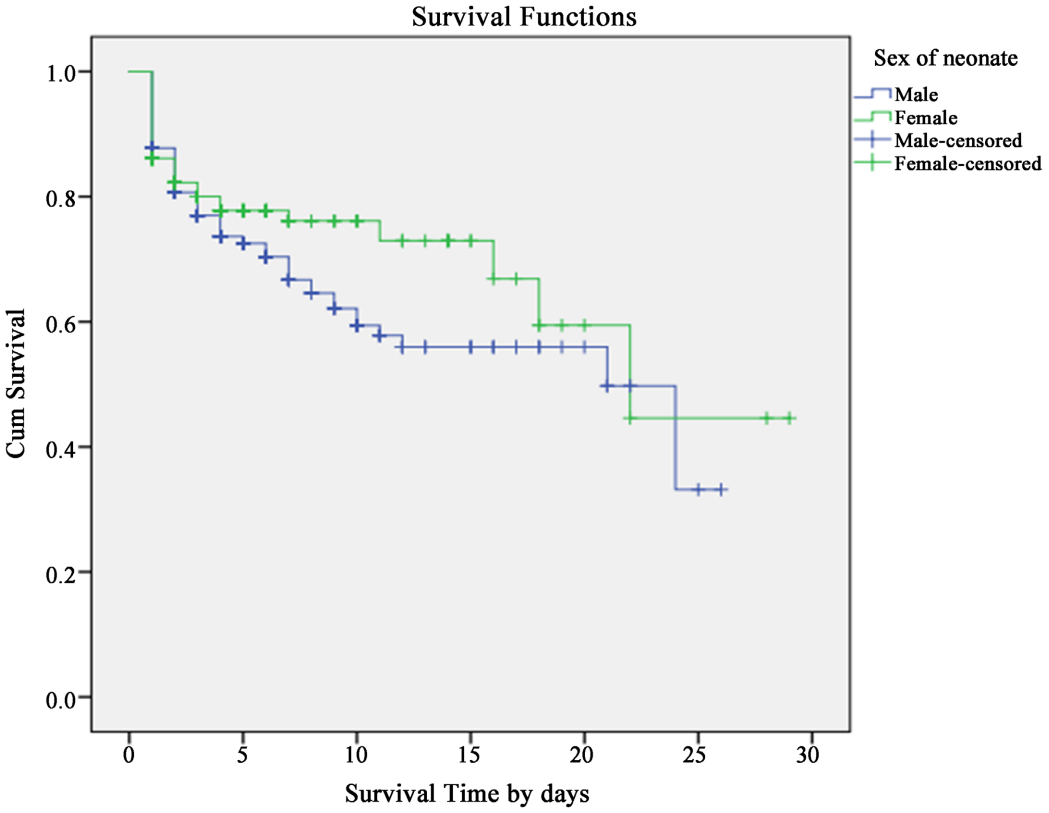


Figure 2. Kaplan-Meier plot or survival curves for the three groups for sex of newborn.

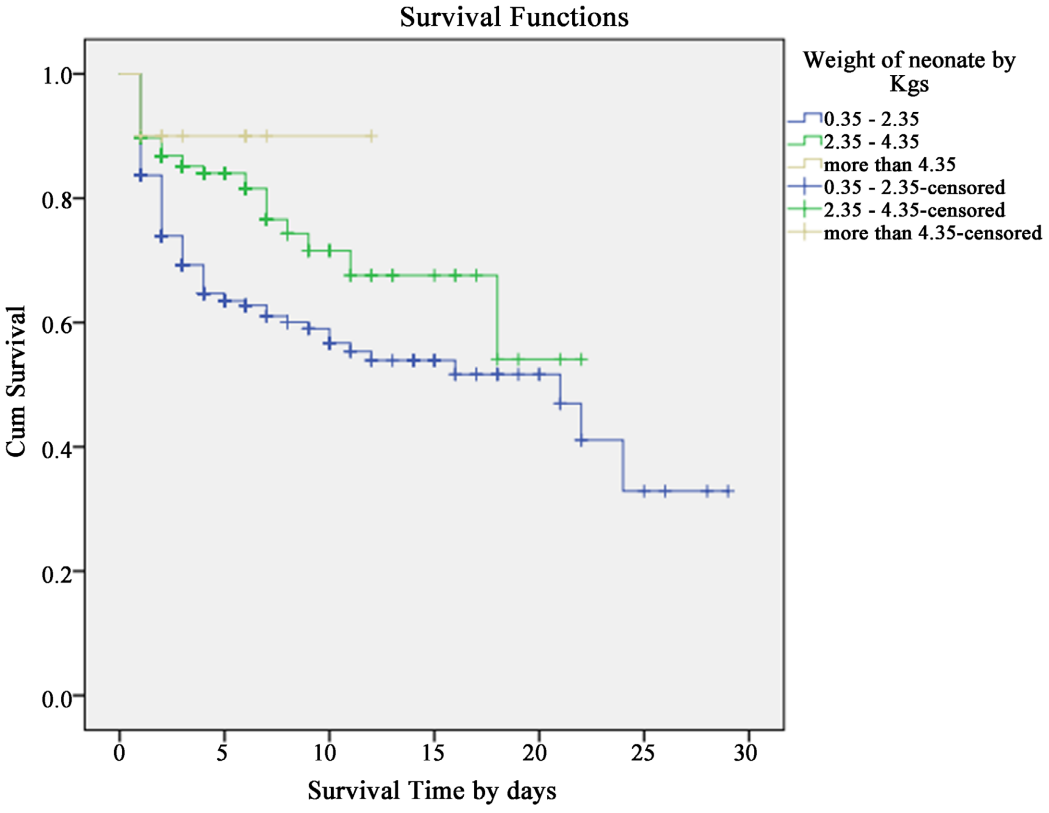


Figure 3. Kaplan-Meier plot or survival curves for three groups of weight of newborn.

than the KM curve for group 1 and group 2. These figures indicate that group 3, which named as (weight of newborn greater than 4.35 kg), has better survival prognosis than group 1, (weight of newborn in period (0.35 - 2.35) kg), and group 2, (weight of newborn in period 2.35 - 4.35 kg).

From Figure 4 we notice that the KM curve for group 3 is consistently higher than the KM curve for group 1 and group 2. These figures indicate that group 3, which named as (gestational age greater than 42 week), has better survival prognosis than group 1, named as (gestational age in period (22 - 32) week) group and group 2, named as (gestational age in period (32 - 42) week).

From Figure 5 we notice that the KM curve for group 1 is consistently higher than the KM curve for group 2. These figures indicate that group 1, which is the normal delivery group, has better survival prognosis than group 2, the caesarian section group.

From Figure 6 we notice that the KM curve for group 1 is consistently higher than the KM curve for group 2. These figures indicate that group 1, which named as urban areas, has better survival prognosis than group 2, rural areas.

From Table 3 we noticed that:

The log-rank statistic's value is (1.86) with P-value (0.173) to three decimal places; as that P-value is greater than 0.05, conclude that there is no significance evidence of a difference in survival times for males and females, that means males and females had approximately same probability and chance to survive.

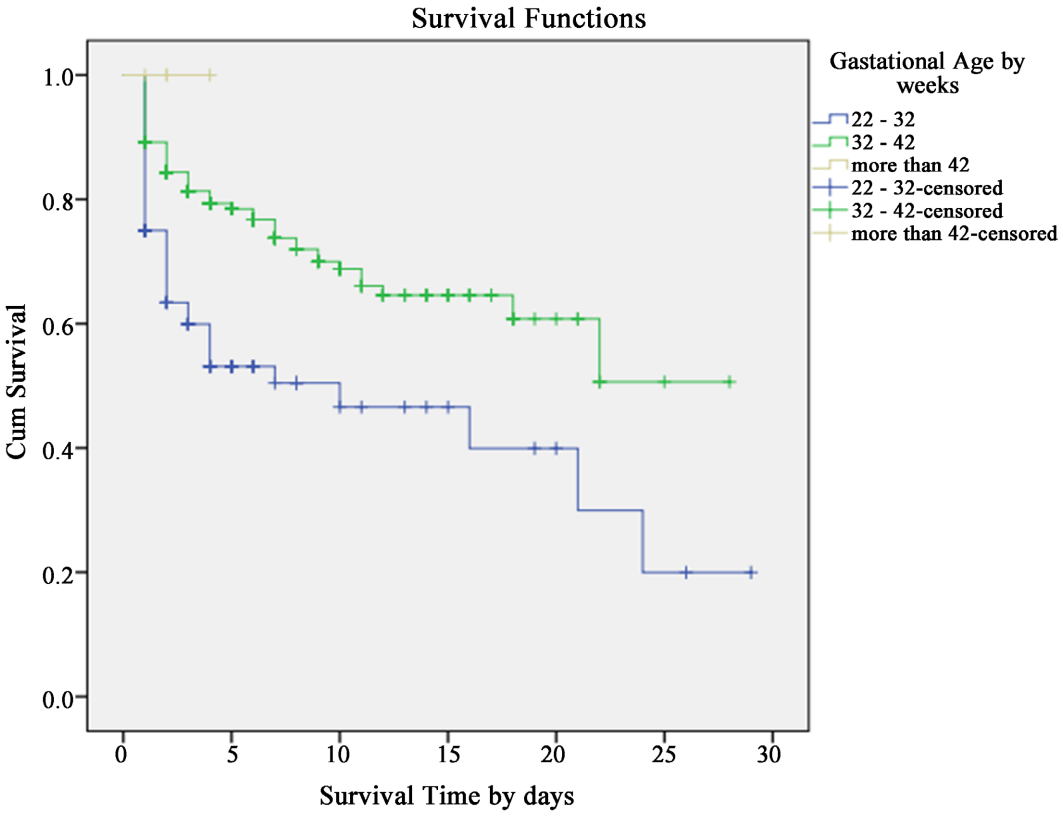


Figure 4. Kaplan-Meier plot or survival curves for three groups of gestational age.

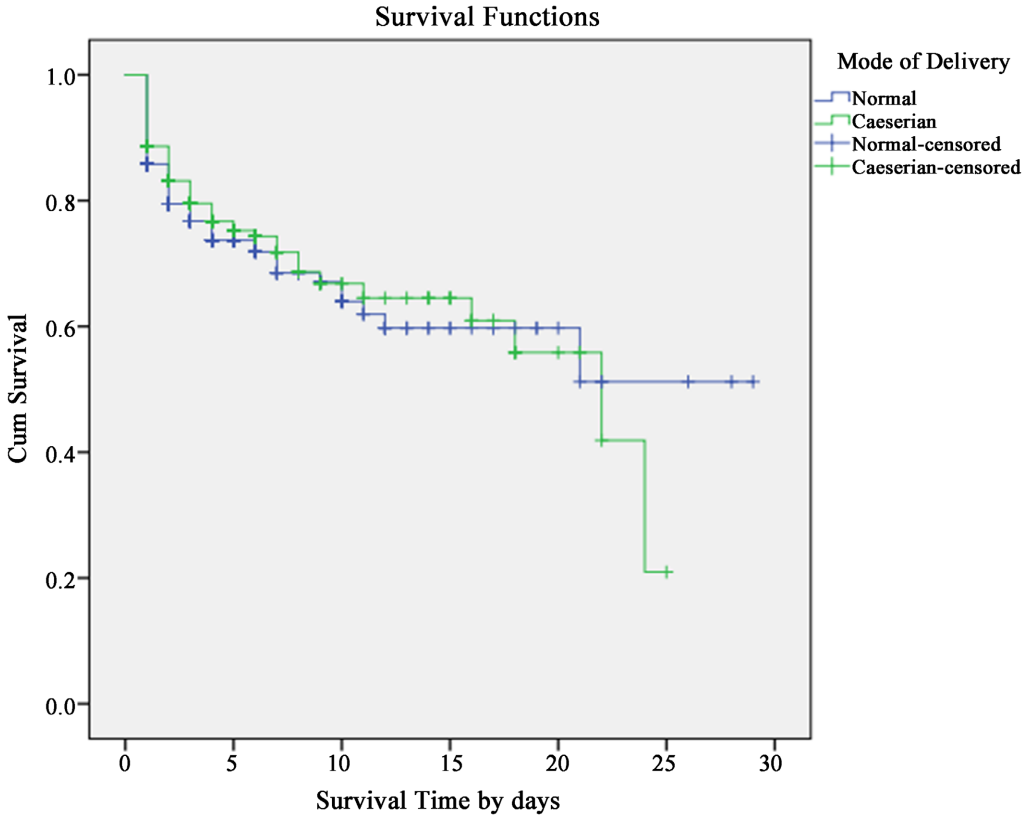


Figure 5. Kaplan-Meier plot or survival curves for the two groups of mode of delivery.

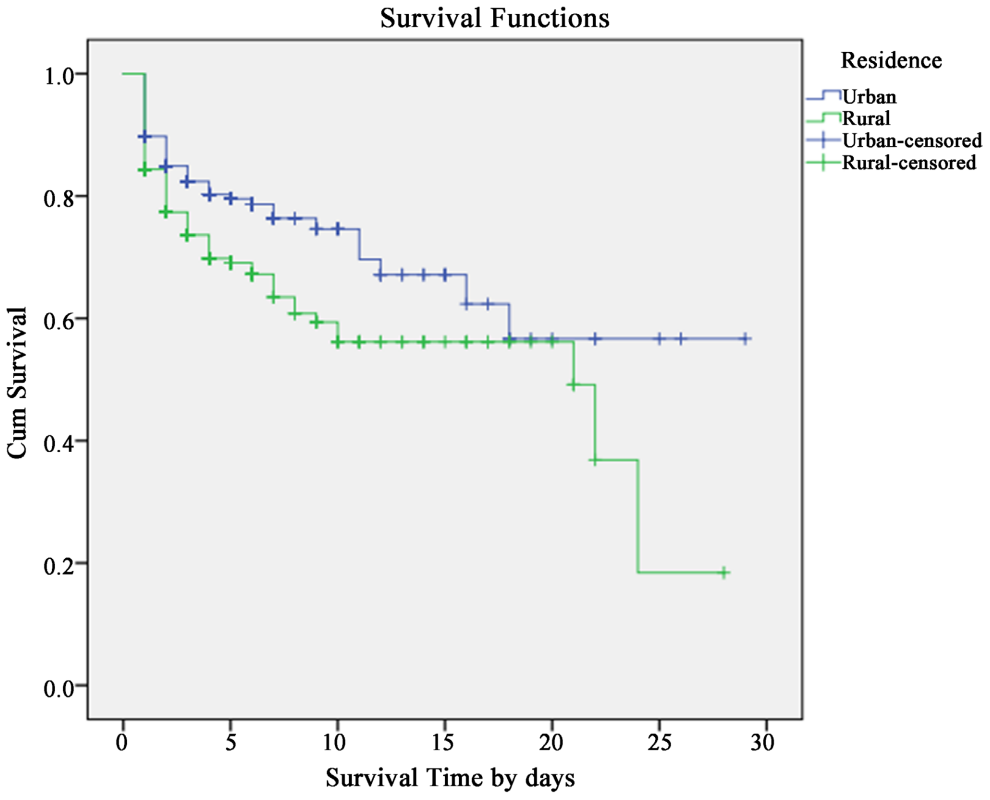


Figure 6. Kaplan-Meier plot or survival curves for two groups of resident type.

Table 3. Summary of the selected factors based on the log-rank test.

Factor	Log-rank statistics	Degree of Freedom	P-value
Sex of newborn	1.86	1	0.173
Weight of newborn	21	2	0.000
Gestational Age	32.51	2	0.000
Mode of Delivery	0.44	1	0.506
Resident Type	9.35	1	0.002

The log-rank statistic is computed to be (21), with P-value zero to three decimal places (0.000); as that P-value is less than (0.05), conclude that there is a high statistically significance difference in survival times between weights of newborn which have been put on scale kg.

The log-rank statistic is computed to be (32.51), with P-value zero to three decimal places (0.000); as that P-value is less than (0.05), conclude that there is a high statistically significance difference in survival times between gestational ages which have been put on scale week.

The log-rank statistic's value is (0.44) with P-value is (0.506) to three decimal places, as that P-value is greater than 0.05, conclude that there is no significant difference in the probability of survival for newborns, whether the delivery is normal or cesarean.

The log-rank statistic's is (9.35) with P-value is (0.002) to three decimal places, as that P-value is less than 0.05, conclude that there is significance evidence of a difference in survival times for urban and rural, so, we can say newborns are significantly more likely to survive in urban areas than in rural areas.

4. Conclusions

According to variable of interest related to the characteristics of a newborn admitted to NICU of hospitals of River Nile State-Sudan, during 2018-2020. We conclude that the gender of a newborn and Mode of Delivery is not significantly different in probability of survival. Whether the gender is male or female, and the delivery is normal or caesarian, weight of newborn, Gestational age and Resident type, have significantly different in probability of survival, which leads us to say the probability of survival increases when the weight of a newborn becomes more than 4.35 kg, Gestational age being greater than 42 weeks at delivery time and when the family of newborn lives in urban areas. The logic of this paper is very strong, and it urges the health sector to increase its ability to control with death of newborns.

We propose some future research topics as follows:

1) The Kaplan-Meier estimates of mortality among newborns admitted with low gestational age at hospitals in Sudan.

2) Log-rank test for comparing neonatal mortality at Governmental Hospitals and Private Hospitals in Sudan.

5. Recommendation

1) Build a new Intensive Care Unit in the study area, especially rural areas.

2) Do other studies more accurate and modern on neonatal mortality, for whole developing countries.

3) Conducting health education convoys for rural areas in Sudan and developing countries.

4) Provide full health care for mother and newborn.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Gomella, T.L., Eyal, F.G. and Mohammed, F.B. (2020) Neonatology: Management, Procedures, On-Call Problems, Disease and Drugs. 8th Edition, McGraw-Hill Education, Alabama.

[2] World Health Organization (2006) Neonatal and Perinatal Mortality: Country, Regional and Global Estimates. World Health Organization, Geneva. https://apps.who.int/iris/handle/10665/43444

[3] Cloherty, J.P., *et al.* (2012) Manual of Neonatal Care. 7th Edition, Lippincott Williams & Wilkins, Philadelphia, 111, 846.

[4] Marschner, I.C. (2005) Inference Principles for Biostatistics. 2nd Edition, Chapman & Hall, New York, 119.

[5] Nikulin, M. and Wu, H.-D.I. (2016) The Cox Model and Its Applications. Springer Nature, Berlin, 25. https://doi.org/10.1007/978-3-662-49332-8

[6] Kleinbaum, D.G. and Klein, M. (2005) Survival Analysis: A Self-Learning Text. 2nd Edition, Springer Science + Business Media, New York, 56-63. https://doi.org/10.1007/0-387-29150-4

[7] Nag, A. (2022) Survival Analysis with Python. Taylor & Francis Group, Abingdon-on-Thames, 3-11.

[8] Kallen, A. (2011) Understanding Biostatistics. John Wiley & Sons, Ltd., Hoboken, 291. https://doi.org/10.1002/9781119992677

[9] Le, C.T. (2003) Introductory Biostatistics. 2nd Edition, John Wiley & Sons, Inc., Hoboken, 381. https://doi.org/10.1002/0471308889

[10] Hanagal, D.D. (2019) Modeling Survival Data Using Frailty Models. 2nd Edition, Springer Nature, East Singapore, 12. https://doi.org/10.1007/978-981-15-1181-3

[11] Hosmer, D.W., *et al.* (2008) Applied Survival Analysis: Regression Modeling of Time-to-Event Data. 2nd Edition, John Wiley & Sons, Inc., Hoboken, 17-44. https://doi.org/10.1002/9780470258019

[12] Rosner, B. (2000) Fundamentals of Biostatistics. 5th Edition, Brooks/Cole, Pacific Grove, 716.

[13] Collett, D. (2003) Modeling Survival Data in Medical Research: Texts in Statistical Science. 2nd Edition, Chapman & Hall, London, 37-48.

[14] Kim, J.S. and Dailey, R.J. (2008) Biostatistics for Oral Health Care. 2nd Edition, Blackwell, Victoria, 289. https://doi.org/10.1002/9780470388303

[15] Chernick, M.R. and Friis, R.H. (2003) Introductory Biostatistics for the Health Sciences: Modern Applications Including Bootstrap. 2nd Edition, John Wiley & Sons, Inc., Hoboken, 343. https://doi.org/10.1002/0471458716