International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Estimate the Best Parametric Survival Models to Determine the Most Important Factors Affecting on Neonatal Mortality in Sudan

Marwa Yousif Ahmed[1], Afra Hashim[2], Elsamoual Mohammed Kurtokeila[3*]
[1,2,3]Department of Statistics, College of science Sudan University of Science and Technology - Sudan
*elsamoual02@gmail.com

***Abstract*— *This study aimed to identify the best model of parametric survival models based on the variables that have a significant effect on neonatal mortality in Sudan. 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, region, 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 through Chi 2 test the parametric models (Log Logistic, Log Normal, Weibull, Exponential and Gompertz) are significant. According to AIC and BIC criterion the Weibull's is the best parametric model.*

***Keywords*— *Survival Model, Neonate, Weibull, Gompertz, Exponential.*

I. INTRODUCTION

Due to the importance of the topic of survival time and its impact on multiple factors, the urgent need has emerged to develop statistical methods and means to increase accuracy and comprehensive and broad knowledge of the factors affecting the survival of the child alive or dead within the study period. Therefore, there are many survival models that deal with this type of studies, including parametric (Exponential Model, weibull Model, log-Logistic Model, logNormal Model, Gompertz Model) and The semi-parametric (Cox Model) These methods are related to the analysis of survival data, which varies according to the nature of the phenomenon studied, and given the importance of these models in practical life and the importance of the child and his future role in the development of nations, many researchers have studied them from many different points of view, and the interest of researchers in these models is still present. In this paper we want to evaluate the most important explanatory variables and using comparison criteria to determine the best parametric model or optimization of parametric survival models identifies the most important factors affecting neonatal mortality in Sudan.

II. DATA & METHODS

Data Collection: Data were collected from Omdurman Maternity Hospital in 2020, 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, region, number of neonate (twin - single) and the sex of neonate (twin - single) and health variables: neonate weight (normal (2.5 - 4) kg, Abnormal) and mode of delivery (normal vaginal delivery {NVD}, cesarean section {C/S}).

Parametric Survival Models: Parametric survival models used in the study, Exponential Model, weibull Model, log-Logistic Model, logNormal Model, Gompertz Model.

Weibull and Exponential Models: Exponential and Weibull models, the proportional-hazards metric simply because it eases comparison with those results produced by stcox (). You can, however, specify the time option to choose the accelerated failure-time parameterization. The Weibull hazard and survivor functions are:

$$h(t) = p\lambda t^{p-1} \quad (1)$$

$$S(t) = \exp(-\lambda t^{p}) \quad (2)$$

Where λ is parameterized ($\lambda_j = \exp(x_j\beta)$) (Peto, 973). If p = 1, these functions reduce to those of the exponential.

Lognormal and Log-logistic Models: For the lognormal distribution, the natural logarithm of time follows a normal distribution; for the log-logistic distribution, the natural logarithm of time follows a logistic distribution. The lognormal survivor and hazard functions are:

$$S(t) = 1 - \Phi\left\{\frac{\log(t)-\mu}{\sigma}\right\} \quad (3)$$

$$h(t) = \frac{{}^{1}/_{\sigma t}\ exp\left[-\frac{1}{2\sigma^2}(\log(t)-\mu)^2\right]}{1-\emptyset\left(\frac{\log(t)-\mu}{\sigma}\right)} \quad (4)$$

Where Φ(z) is the standard normal cumulative distribution function.

The lognormal regression is implemented by setting $\mu_j = x_j\beta$ and treating the standard deviation, σ, as an ancillary parameter to be estimated from the data.

The loglogistic regression is obtained if zj has a logistic density. The log-logistic survivor and hazard functions are:

$$S(t) = \left\{1 + (\lambda t)^{1/\gamma}\right\}^{-1} \quad (5)$$

$$h(t) = \frac{\lambda\gamma t^{\gamma-1}}{1+\lambda t^{\gamma}} \quad (6)$$

This model is implemented by parameterizing $\lambda_j = \exp(-x_j\beta)$ and treating the scale parameter γ as an ancillary parameter to be estimated from the data.

Gompertz Model: The Gompertz regression is parameterized only as a PH model. This model has been extensively used by medical researchers and biologists modeling mortality data. The Gompertz distribution implemented is the two-parameter function as described in (Lee and Wang, 2003), with the following hazard and survivor functions:

Marwa Yousif Ahmed, Afra Hashim, and Elsamoual Mohammed Kurtokeila, “Estimate the Best Parametric Survival Models to Determine the Most Important Factors Affecting on Neonatal Mortality in Sudan,” *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 5, Issue 7, pp. 99-102, 2023.

h(t) = λ exp (γt) (7)

S(t) = $exp\{-\lambda\gamma^{-1}(e^{\gamma t}-1)\}$ (8)

The model is implemented by parameterizing $\lambda_j = exp(x_j\beta)$, where γ is an ancillary parameter to be estimated from the data.

Criteria of evaluation:

I. *Akaike Information Criterion (AIC):* Standard AIC was proposed by (AKAIKE -1974) to measure the quality and accuracy of the statistical model, and it is given by the following formula:

AIC= -2logL+2k (9)

number of parameters ≡ k

Likelihood ≡ L

II. *Bayesian Information Criterion (BIC):* Standard BIC was proposed by (Scnwarz's, 1978) to measure the quality and accuracy of the statistical model estimation, taking into account the sample size:

BIC=-2logL+klogN (10)

III. Results

Test the significant factors:

TABLE 1: Estimating parameters of parametric survival models:

Model \ Factors		Age	Previous delivers	Sex neonate	Number of neonate	City	Weight neonate	Mode of delivery
Exponential	parameter	0.692	0.154	0.778	1.106	1.026	2.84	0.63
	Wald test	2.03	2.99	1.25	0.01	0.99	3.93	0.73
	P-value	0.01	0.00	0.21	0.99	0.32	0.00	0.47
Weibull	parameter	0.849	0.404	0.820	1.778	3.40	4.474	0.033
	Wald test	1.08	2.23	1.26	0.01	0.24	4.81	0.04
	P-value	0.04	0.01	0.21	0.99	0.15	0.00	0.97
Logistic	parameter	0.334	0.777	0.335	2,799	1.00	1.408	0.248
	Wald test	1.94	3.4	1.82	0.01	1.31	5.7	0.94
	P-value	0.05	0.02	0.06	0.99	0.08	0.00	0.35
Lognormal	parameter	0.335	0.077	0.335	2,799	1.004	1.408	0.248
	Wald test	1.94	3.31	0.34	0.01	1.82	5.7	0.94
	P-value	0.05	0.00	0.74	0.99	0.07	0.00	0.35
Gompertz	parameter	1.214	0.569	1.214	0.870	4.254	1.717	0.026
	Wald test	3.47	2.6	1.74	0.00	0.03	4.39	0.00
	P-value	0.00	0.01	0.08	1.00	0.98	0.00	1.00

Source: prepared by the researchers by using STATA 17, 2022

TABLE 2: Log rank test for parametric survival models

Models	Log rank test	Age	Previous delivers	Sex neonate	Number of neonate	City	Weight neonate	Mode of delivery
Exponential	Chi^2	12.42	12.02	0.02	3.42	0.23	18.41	0.48
	p-value	0.01	0.03	0.62	0.06	0.63	0.00	0.49
Weibull	Chi^2	14.25	11.1	0.31	3.02	0.15	19.36	0.97
	p-value	0.01	0.03	0.58	0.08	0.70	0.00	0.32
Log-logistic	Chi^2	21.8	21.03	0.41	14.4	0.00	22.2	1.09
	p-value	0.01	0.05	0.522	0.08	0.99	0.00	0.30
Lognormal	Chi^2	20.34	16.96	0.43	3.16	0.03	20.33	1.1
	p-value	0.00	0.01	0.51	0.08	0.86	0.00	0.30
Gompertz	Chi^2	10.79	18.8	0.40	2.90	0.13	18.67	1.15
	p-value	0.02	0.01	0.53	0.08	0.72	0.00	0.28

Source: prepared by the researchers by using STATA 17, 2022

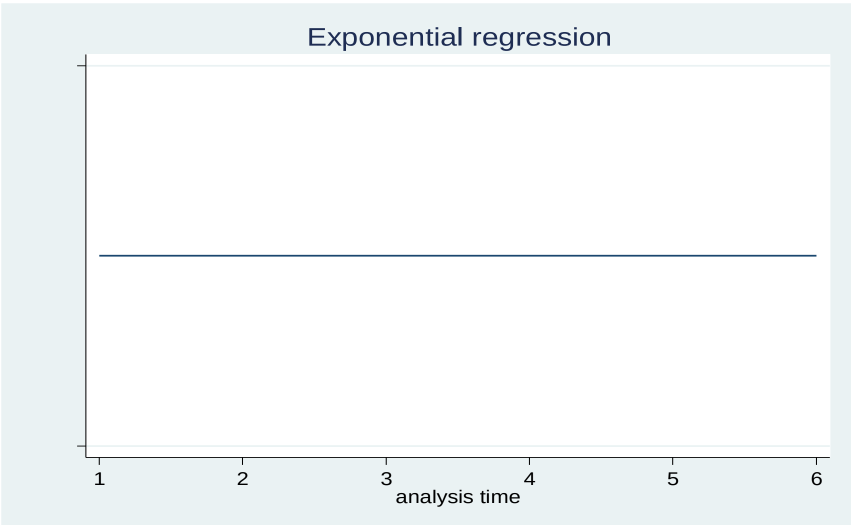

Figure 1: The hazard function of the Exponential model

Source: prepared by the researcher by using STATA 17, 2022

Marwa Yousif Ahmed, Afra Hashim, and Elsamoual Mohammed Kurtokeila, "Estimate the Best Parametric Survival Models to Determine the Most Important Factors Affecting on Neonatal Mortality in Sudan," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 5, Issue 7, pp. 99-102, 2023.

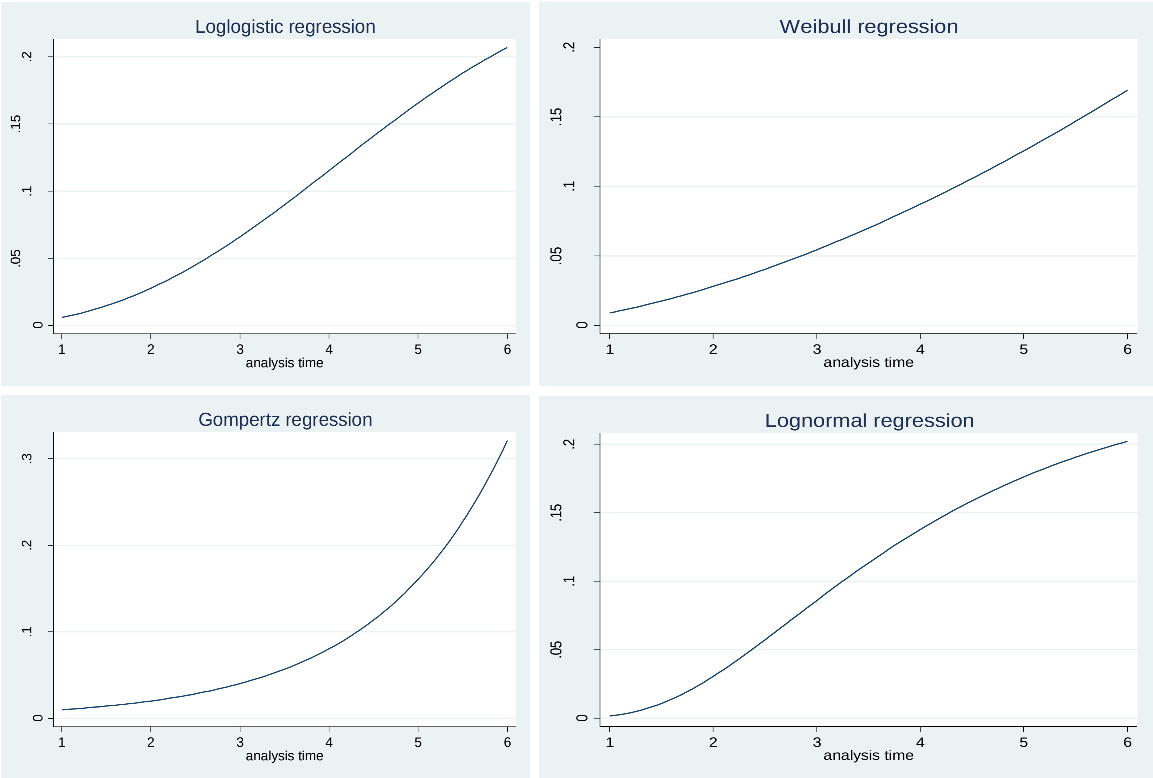

Figure 2: Estimated Hazard Function for Parametric Survival Models:
Source: prepared by the researchers by using STATA 17, 2022

TABLE 3: Estimation of parameters of significant variables in parametric survival models:

Parameter	Exponential	Weibull	Log-Logistic	Lognormal	Gompertz
Age	0.227	0.37	0.09	0.11	0.42
Parous	0.611	-1.71	0.45	0.38	1.77
Weight neonate	2.88	3.74	1.38	1.33	3.70
constant	5.48	7.15	2.79	2.77	6.54
P	-	2.64	-	-	-
Γ	-	-	0.31	-	0.69
σ	-	-	-	0.55	-

Source: prepared by the researchers by using STATA 17, 2022

TABLE 4: Criteria for choosing the best model:

Models	Log likely	Chi^2	P-value	AIC	BIC
Exponential	20.20	19.42	0.00	42.40	42.13
Weibull	14.49	24.00	0.00	32.98	32.45
Logistic	15.21	22.82	0.00	34.42	33.88
Lognormal	15.03	22.94	0.00	34.06	33.52
Gompertz	15.30	23.30	0.00	34.60	34.06

Source: prepared by the researchers by using STATA 17, 2022

IV. DISCUSSION

From table 1: We found that the factors of (previous delivers and weight neonate) had significant effect on neonatal mortality at the level of significance 5% for all study models, and the factor of mother's age is significant in models of (Exponential, Weibull and Gompertz). But the factors of (sex neonate, number of neonate, city and mode of delivery) are not significant for all study models; this means that there had not effect on neonatal mortality at the level of significance 5%.

From table 2: We found that the factors of (mother's age previous delivers and weight neonate) had significant differences in hazard probabilities through survival time at the level of significance 5% for all study models except factor of previous delivers in Log-logistic model and the factors of (sex neonate, number of neonate, city and mode of delivery) had not significant differences in hazard probabilities through survival time at the level of significance 5%.

The hazard function of the exponential model is a constant function with time and increased shown in figures 1 & 2.

Table 3 shows that the estimation of the parametric survival models for the significant factors (age, previous delivery, weight neonate) and the estimation of the parameter of each model.

From table 4: We found that the probabilities values of all parametric survival models (0.00<0.05) means that all models are significant and can be used in estimating the hazard of neonatal mortality, the Weibull model has the lowest criterion values of (AIC and BIC) with chi^2 (24); The best parametric survival model for estimating the hazard of death for neonatal mortality is Weibull and the model given by the following formula (referring to Table 3):

$$\hat{h}(t_j|x_j) = \{\exp(7.15 + 0.37Age_j - 1.71parous_j + 3.74weight\ neonate_j\}\ 2.64t_j^{1.64}$$

REFERENCES

1. Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723.
2. Cleves, M. A., W. W. Gould, R. G. Gutierrez, and Y. V. Marchenko. 2010. An Introduction to Survival Analysis Using Stata. 3rd ed. College Station, TX: Stata Press.
3. Cox, D. R., and D. Oakes. 1984. Analysis of Survival Data. London: Chapman & Hall/CRC.
4. Crowther, M. J., K. R. Abrams, and P. C. Lambert. 2013. Joint modeling of longitudinal and survival data. Stata Journal 13: 165–184.
5. Gutierrez, R. G. 2002. Parametric frailty and shared frailty survival models. Stata Journal 2: 22–44.
6. Hosmer, D. W., Jr., S. A. Lemeshow, and S. May. 2008. Applied Survival Analysis: Regression Modeling of Time to Event Data. 2nd ed. New York: Wiley.
7. Hougaard, P. 1986. Survival models for heterogeneous populations derived from stable distributions. Biometrika 73:387–396.
8. Kalbfleisch, J. D., and R. L. Prentice. 2002. The Statistical Analysis of Failure Time Data. 2nd ed. New York: Wiley.
9. Klein, J. P., and M. L. Moeschberger. 2003. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York: Springer.
10. Lambert, P. C., and P. Royston. 2009. Further development of flexible parametric models for survival analysis. Stata Journal 9: 265–290.
11. Lee, E. T., and J. W. Wang. 2003. Statistical Methods for Survival Data Analysis. 3rd ed. New York: Wiley.
12. Peto, R., and P. Lee. 1973. Weibull distributions for continuous-carcinogenesis experiments. Biometrics 29: 457–470.
13. Royston, P., and P. C. Lambert. 2011. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model.College Station, TX: Stata Press.
14. Royston, P., and M. K. B. Parmar. 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: 2175–2197.
15. Weibull, W. 1939. A statistical theory of the strength of materials. In Ingeniors Vetenskaps Akademien Handlingarvol. 151. Stockholm: Generalstabens Litografiska Anstalts Forlag.