Recurrent Malaria In Kenya: Survival Distribution And Equality Of Survival Distribution

Article · December 2017

CITATIONS
0

READS
80

5 authors, including:

Erick Okuto
Jaramogi Oginga Odinga University of Science and Technology
26 PUBLICATIONS **86** CITATIONS
SEE PROFILE

Dickens Aduda
University of Kabianga
11 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Charles O Obonyo
Kenya Medical Research Institute
55 PUBLICATIONS **1,879** CITATIONS
SEE PROFILE

Recurrent Malaria In Kenya: Survival Distribution And Equality Of Survival Distribution

Ondulo James
Jaramogi Oginga Odinga University of Science and Technology, Kenya

Okuto Erick
Jaramogi Oginga Odinga University of Science and Technology, Kenya

Muga Maua
Jaramogi Oginga Odinga University of Science and Technology, Kenya

Omondi Dickens
University of Kabianga, Kenya

Obonyo Charles
Kenya Medical Research Institute, Centers for disease control and prevention, Kisumu, Kenya

ABSTRACT

Malaria is a leading cause of morbidity and mortality in Kenya, mostly affecting the rural poor, especially young children and pregnant women. Most clinical studies in malaria involve a number of equally important endpoints. This would normally portend challenges in relation to issues about the study design, analysis of the data and interpretation of the results. The majority of malaria related clinical studies do not factor in or experience difficulties in analyzing recurrent events. It is desirable to utilize multiple event times in the analysis procedures to obtain efficient inferences for therapeutic effect over time and to account for the dependence of the recurrent events in an individual. In this paper we aimed to determine the parametric distributions as an approximation of the KM nonparametric distribution of the recurrent malaria data. We considered survival distribution and the equality of survival distribution in analyzing specific covariates (gender, anemia and drug treatments) in a recurrent malaria data. We used *R* software to construct Kaplan – Meier survival curves and demonstrate their equality by running a Log- rank test to determine their performance and find the level of significance of the covariates survival distribution. The result showed that there was no significant difference at 95% Cl of the covariates under scrutiny. The significance level for the covariates were: gender (p-value= 0.41); anemia (p-value= 0.816); the treatment drug combination 2 (p-value= 0.637) and treatment drug combination 1 (p-value= 0.808). However, majority of the patients who had recurrent malaria also had anemia. Understanding the survival distribution and the equality of the survival distribution in recurrent malaria cases is essential for designing optimal statistical procedures that do not bias study results.

Keywords: Recurrent events, Malaria, Kaplan- Meier, Log- rank test, survival distribution, survival function, hazard function, parametric distributions, nonparametric distributions, Akaike Information Criteria (AIC), Log-likelihood function.

1. INTRODUCTION

In a follow-up clinical trial to evaluate a new treatment, quite often each study subject may experience a number of 'failures' that correspond to repeated occurrences of the same type of event or events of entirely different natures during his/her follow-up period. This is particularly true when there are a number of equally important endpoints involved in the trial. In malaria related clinical studies, the subject may experience a number of treatment failures which correspond to repeated occurrences of the same type of infection (recrudescence) or to the occurrence of events of entirely different natures (new infections)(Checchi et al., 2005; Obonyo, 2006). Recurrences may be due to heterogeneity across individuals associated with unknown, unmeasured, or unmeasurable effects or due to occurrence dependence associated with antecedent event(s) that may make further events more or less likely (Box-Steffensmeier & De Boef, 2006). Analysis of recurrent events require special statistical approaches because of censoring, which is defined as not knowing the exact time of the

survival event. Of interest is mostly the hazard and survival functions. It is desirable to utilize multiple event times in the analysis to obtain efficient inferences for the therapeutic effect over time, and to account for the dependence of the recurrent events in an individual and any unobserved heterogeneity of the event propensity across individuals.

Inferences on recurrent malaria have critical implications for policy and practice guidelines. However, multiple failure time data normally portend challenges in relation to issues about the study design, analysis of the data and interpretation of the results (Hagar & Dukic, n.d.). In the current study survival analysis approach was used to assess recurrence after initial treatment upon positive diagnosis that a child has malaria. One of the common uses of survival analysis in clinical trials is the comparison of survival times given different treatment and different response or survival of subjects to time in the event under consideration, in this case, occurrence/recurrence of malaria. In this article, the specific interest is to determine the impact of anemia, gender and the treatment drugs on malaria recurrence and their respective survival based on malaria data from a clinical randomized malaria treatment trial in a pediatric population. We used Kaplan Meier which is a non-parametric method to estimate the survival distributions across the various covariates and the Log-rank test to determine the equality of these survival distributions. We also assessed the performance of the model distribution.

1.1 SURVIVAL DISTRIBUTION: TYPES AND THEIR CHARACTERISTICS

Various survival distributions exist. They can be parametric or non-parametric distributions. The most commonly used one is non-parametric distribution is Kaplan Meier (KM). The commonly used parametric distributions are: exponential, Weibull, logistic, log-logistic, Gaussian, log- normal, Rayleigh, and Student-t (Butler, 2011). Each of these distributions have different characteristics which make them suitable for different data sets depending on their probability densities. Based on given parametric curves we can: *i)* derive concise equations and smooth functions for estimating the survival function - S(t), cumulative hazard - H(t) and hazard function - h(t); *ii)* estimate the survival function more precisely than KM assuming the parametric form is correct and also estimate the expected failure time ("Lecture 15 Introduction to Survival Analysis," 2004). Sometimes we may want to make more assumptions that allow us to model the data in more details. Here below is an example: exponential distribution which has a probability density,

$$p(t) = \lambda e^{-\lambda t} \qquad \text{(i)}$$

The exponential distribution becomes, $P(t) = \int_0^t p(x)dx$

$$= 1 - e^{-\lambda t} \qquad \text{(ii)}$$

The survival function then becomes, S(t) = 1- P(t)

$$= 1-\{1- e^{-\lambda t}\}$$

$$= e^{-\lambda t} \quad \text{(iii)}$$

The hazard function is, $h(t) = \frac{p(t)}{S(t)}$, $= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}$

$$\frac{-d}{dt} \log S(t) = \lambda \quad \text{(iv)}$$

And the Cumulative hazard becomes, $H(t) = \int_0^t h(x)dx$

$$= \log_e S(t)$$

$$= \lambda t \quad \text{(v)}$$

The exponential distribution is helpful in interpreting the meaning of the hazard rate but is not practicable in biological processes where the hazard rate is not constant (Lee & Go, 1997). Hence use of other probability distributions which give variable hazard rates to model the survival data is recommended, for example Gompertz and Weibull distributions. These two specifically reduce to exponential when the hazard rate is constant.

Here below is the summary table of a few survival distributions and their corresponding densities, survival function, and hazard rate.

Table 1.1: The Parametric Distributions and their respective: densities, survival functions and the hazard functions

Distribution	Parameter	Density, f(t)	Survival function, S(t)	Hazard function, h(t)
Exponential	$\lambda > 0$	$f(t) = \lambda e^{-(\lambda t)}$	$e^{-(\lambda t)}$	$h(t) = \lambda$
Lognormal	$\mu, \sigma > 0$ $a = \exp(-\mu)$	$G(y) = \frac{1}{\sqrt{2\pi}} \int_0^y \exp(-u^2/2)\, du$ $f(t) = 1/t\sigma\sqrt{2\pi}\ exp[-1/2\sigma^2\ (lnat)^2]$	S(t) = $1 - G(ln\ at/\sigma)$	$h(t) = \frac{1/t\sigma\sqrt{(2\pi)exp\left[-1/2\sigma^2(\log_e at)^2\right]}}{1 - G(\log_e(at/\sigma))}$
Log-logistic	$\lambda, \quad \gamma > 0$	$f(t) = \lambda\gamma(\lambda t)^{\gamma-1}[1 + (\lambda t)^{\gamma}]^{-2}$	$S(t) = [1 + (\lambda t)^{\gamma}]^{-1}$	$h(t) = \lambda\gamma(\lambda t)^{\gamma-1}[1 + (\lambda t)^{\gamma}]^{-1}$
Weibull	$\lambda, \quad \gamma > 0$	$f(t) = \lambda\gamma(\lambda t)^{\gamma-1}exp(-\lambda t)^{\gamma}$	$S(t) = exp(-\lambda t)^{\gamma}$	$h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$
Gompertz	$\beta > 0, \gamma > 0$	$f(t) = \beta e^{\gamma t - \frac{\beta}{\gamma}(e^{\gamma t}-1)}$	$S(t) = e^{-\frac{\beta}{\gamma}(e^{\gamma t}-1)}$	$h(t) = \beta e^{\gamma t}$
Gompertz-Makeham	$\beta > 0, \gamma > 0, \alpha > 0$	$f(t) = (\alpha + \beta e^{\gamma t}) e^{-\alpha t - \frac{\beta}{\gamma}(e^{\gamma t}-1)}$	$S(t) = e^{-\alpha t - \frac{\beta}{\gamma}(e^{\gamma t}-1)}$	$h(t) = \alpha + \beta e^{\gamma t}$

As some information may be lost due to censoring other methods more effective than the common KM have been suggested to minimize misclassification bias. For example, we can use prognostic covariate information to recover some of the information lost due to censoring (Lee & Go, 1997). Parametric approach provides an additional advantage in identifying prognostic or risk factors. Parametric distributions also have the advantage of taking care of the inherent measurement errors in longitudinal covariates, which tend to bias the association parameters (Ozg, 2015). Consequently, they minimize biased predictions due to over- or under estimations of study parameters. Whereas the nonparametric analysis of survival distribution takes care of the real data as it is without exaggeration, its limitation is that it is not an effective tool in relation to forecasting and policy making. Parametric distribution assumes a homogeneous scenario which may not be the case at all times (Wienke, 2003).

Therefore, it is necessary to correctly choose the right distribution to ensure that the resulting model is more robust to support decision making arising from them.

1.2 IMPLICATIONS OF IDENTIFYING DISTRIBUTION TYPES IN PERFORMING SURVIVAL ANALYSIS

The model and covariates in censored data as well as of the estimators used in such data exhibit unknown distribution (Su, 2016) while the parametric methods assume known distribution and that survival times of all subjects are exact and known (Lee & Go, 1997) Additionally, there are several techniques of multivariate analyses. Depending on the type of outcome variable, different techniques are used; if the outcome is a continuous variable, we use linear regression; if the outcome is categorical, we use logistics regression; if the outcome is 'time to event', we use Cox proportional hazard model and if the outcome is 'number of events' (discrete numeric), then we use Poisson regression (Multivariable analysis in R part 2 : Cox proportional hazard model," n.d.). The outcome measure of interest for the current paper is 'time to event', and hence we extend the use of Kaplan- Meier analysis for the univariate analysis and Cox proportional hazard model for multivariate analysis, of recurrent malaria occurring in children less than 60 months old who had participated in a randomized controlled malaria treatment study.

To investigate the underlying parametric distribution of the data, one can fit various survival functions to the data and compare how similar the survival functions are to the Kaplan- Meier estimates of the survival function (Butler, 2011). In the case of this data, it is hypothesized that the data fits best a Student's *t*-distribution. This is arrived at after running R program to determine the outcome of the various probability distributions and using AIC to select the most suitable distribution. Graphing the survival function is important because it provides valuable insight into the behavior of the data (Su, 2016). The Kaplan-Meier estimate of the survival curve will remain flat until there is a failure at which time the curve will drop, an amount proportional to how many items/ people failed (experienced the event) at that time. When there are no more failures, the curve will flatten out again until it reaches another failure, where again the drop of the curve with be proportionate to the amount of children that have failed/ experienced malaria recurrence. The shape of the KM distribution curve in this data therefore fits the t-distribution as already discussed above that it belongs to elliptical distributions(Hofert, 2013). This means that the steeper the curve, the more failures there are and the larger the hazard rate. If the curve is relatively flat and has a shallow slope, there are a lot of children surviving, i.e. children are failing/ experiencing the event (malaria) at a slow rate. The Kaplan-Meier estimate of the survival curve does not depend on any parametric assumptions about the underlying probability distribution of the data (Butler, 2011).

2. METHODS

The secondary data for malaria was analyzed for recurrent malaria episodes among the under 5years in who participated in a randomized clinical trial conducted in Siaya County. AG-CP was used to reconstruct the initial

data and work out recurrent cases. To determine the survival distribution and the equality of survival distribution we adopted the nonparametric survival function, Kaplan- Meier survival function curve, from which other survival distributions and functions can be derived and comparison made. By using R software, the various covariates, anemia1, gender, and treatment drugs (rancode and rancode2) were considered to determine their survival functions and curves. A log-rank test was also done to determine how the covariates significantly influence the survival. Various probability distributions were fitted to the data and a comparison was done to determine the most appropriate parametric survival distribution. The comparison was done using AIC and log-likelihood ratio test. The multivariate normal distribution can be defined in various ways, one is with its stochastic representation

$$X = \mu + AZ \quad (1)$$

The multivariate t distribution with ν degrees of freedom can be defined by the stochastic representation

$$X = \mu + \sqrt{W}AZ, \quad (2)$$

Where $\sqrt{W} = \nu/\chi_\nu^2$ (χ_ν^2 is informally used here to denote a random variable following a chi-squared distribution with $\nu > 0$ degrees of freedom) is independent of Z and all other quantities are as in (1). By introducing the additional random factor √W, the multivariate t distribution with ν degrees of freedom is more flexible than the multivariate normal distribution. As for the multivariate normal distribution, the density (3) has ellipsoidal level sets and thus belongs to the class of elliptical distributions(Hofert, 2013).

Density function for a Student- t distribution is given as:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\cdot\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \text{ for } -\infty < t < \infty \quad (3)$$

The Students *t*-distribution, while useful, is rarely used in describing survival analysis distributions, probably due to the mathematical complexities involved. The *R* software works it out and only gives the output result. It is prudent to note here that Student-*t* tends to standard normal distribution when the degree of freedom gets larger and that it is also bell- shaped like any normal distribution. The *t*-distribution provides a useful alternative of error distribution for modelling of data sets with longer-than-normal tails. It is able to capture and account for the extreme values which the normal distribution does not. This makes the longitudinal data analysis using *t*-distribution more robust than the classical models (Song, 2013). In this study, the Student-*t* distribution appropriately fits the data since in recurrence of malaria and according to the study design, the tail gets longer with fewer patients remaining with recurrent malaria as the majority experience complete clearance when more effective antimalarial drugs are used.

3. RESULTS

3.1 SUMMARY VALUES OF COVARIATES

Table 3: Summary of malaria recurrence by gender, anemia and treatment drug combination

Variable		N	Recurrence
Gender	Male	546	7
	Female	647	10
Anemia	No	101	1
	Yes	1092	16
Rancode	AS1	340	4
	AS3	441	6
	SP	412	7
Rancode2	AS1	781	10
	AS3	412	7

Table 3.1 below shows the *Log- rank* tests done on the data to determine survival distributions equality by specific variable responses and the results of the differences obtained. The log- rank test shows that there is no difference in performance of the survival curves for each of the tested variables (gender, anemia, and treatment-rancodes). Visually this is illustrated by the curves crossing, an indication of inconsistency in performance.

Table 3.1: Log- rank statistics test for equality of survival distribution

Variable		N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$	P- value
Gender	Male	546	7	8.48	0.257	0.678	0.41
	Female	647	10	8.52	0.256	0.678	
Anemia	No	101	1	0.801	0.04943	0.0542	0.816
	Yes	1092	16	16.199	0.00244	0.0542	
Rancode	AS1	340	4	3.7	0.0241	0.346	0.808
	AS3	441	6	7.13	0.1782	0.4120	
	SP	412	7	6.17	0.1111	0.2330	
Rancode2	AS1	781	10	10.83	0.0633	0.233	0.637
	AS3	412	7	6.7	0.1111	0.233	

3.2 KAPLAN MEIER (KM) SURVIVAL CURVES

The KM survival curves for each specific covariates gender (male and female), anemia (present or absent) treatment combination (Rancode - AS1 and AS3; Rancode 2 - AS1, AS3 and SP) revealed that distribution of recurrence of malaria in response to each covariate was not significant as shown by p-values in table 3.0. Figure 3.1A shows that the lines corresponding to AS1 and AS3 cross at various points indicating that the treatment drugs are not consistently effective in preventing recurrence of malaria. The same is confirmed by the insignificant p-value of 0.808 at 95% Cl. Rancode2, in figure 3.1B another randomized treatment drugs has the same trend with the p- value of 0.637 at 95% Cl. Figure 3.1C also reveal inconsistent response among the gender, with the lines for male and female crossing at various points of the whole study time. Figure 3.1D shows that few of the patients who were not anemic experienced recurrence of malaria while the majority of those who were anemic experienced recurrent malaria.

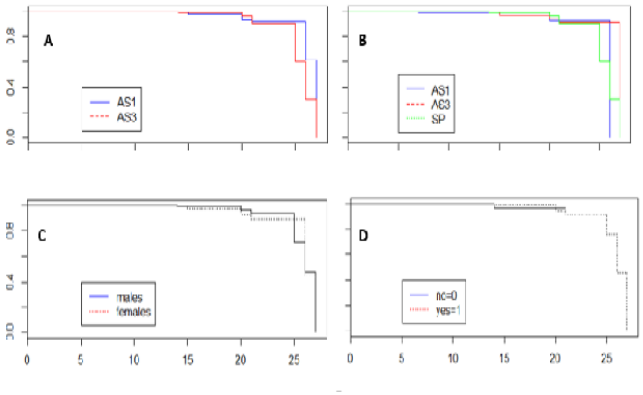

*Figure 3.1: Kaplan- Meier survival curves for covariates: **A** rancode (randomized treatment drug- AS1 and AS3), **B** rancode2 (randomized treatment drugs- AS1, AS3 and SP), **C** gender (males and females) and **D** anemia level (not anemic=0 and anemic=1).*

The KM survival distribution curve shown in figure 3.2 gives a flat top, revealing that recurrence can only occur after the initial occurrence of malaria in a patient and also implies that the rate of survival function is low, not steep gradient. After the first recurrence, the curve then takes the normal KM shape and ends with the stop of the follow up time, which was day 28.

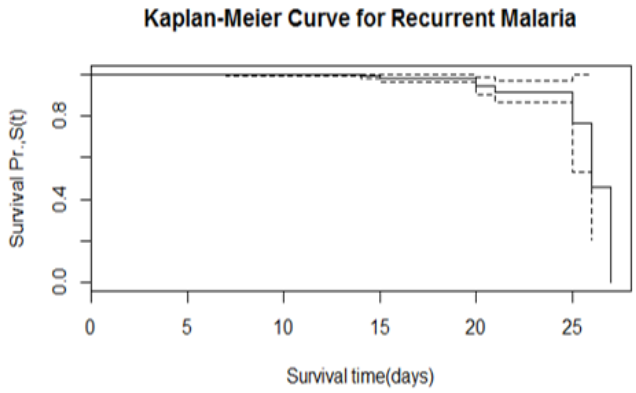

Figure 3.2: The KM Survival Curve for the Recurrent Malaria Data of children under 5years of age in Siaya County

The output below table 3.3 is the result of KM survival estimates. The same can be achieved by using the formula:

$$S(t) = \prod_{i}^{n}\left(1 - \frac{d_i}{n_i}\right)$$

***Table 3.3**: The Survival results of the under 5years children who were treated of recurrent malaria episodes in Siaya County, between 1999 and 2000.*

survfit(formula = Surv(tstart, genotype) ~ 1, data = newsubdf110_T)

Time	n.risk= n_i	n.event=d_i	survival	std.err	lower 95% CI	upper 95% CI
7	613	1	0.998	0.00163	0.995	1.000
14	319	2	0.992	0.00470	0.983	1.000
15	109	1	0.983	0.01019	0.963	1.000
20	103	4	0.945	0.02112	0.904	0.987
21	96	3	0.915	0.02646	0.865	0.969
25	6	1	0.763	0.14099	0.531	1.000
26	5	2	0.458	0.18730	0.205	1.000
27	3	3	0.000	Na	Na	Na

The Summary of the Parametric distributions as ranked by AIC and log- likelihood values, and as fitted in the recurrent malaria data of the under 5years of age children in Siaya County between 1999 and 2000 are shown on table 3.4 below.

Table 3.4: Summary of the Parametric Distributions as ranked by AIC and Loglikelihood

Rank	Distribution	AIC	Loglikelihood
1	Student- t	177.4	-86.7
2	Logistic	179.4	-87.7
3	Weibull	180.7	-88.3
4	Log-logistic	181.6	-88.8
5	Gaussian	182.3	-89.1
6	Log-gaussian	185.6	-90.8
7	Rayleigh	230.6	-114.3
8	Exponential	265.5	-131.8
9	Extreme	-3736.3	1870.1

It shows that the most appropriate parametric distribution for the data is Student-*t-distribution* and that exponential distribution does not rate well for the model. This is true since exponential model would imply a steep gradient unlike what the KM graph gives.

The table 3.5 below gives the results of the survival distributions for the specific covariates and their respective AIC values, the *p*- values and the log-likelihood values of the distribution. The *t*-distribution is the best fitted survival distribution for this data, based on the AIC values. It is ranked first for all the covariates (gender; Rancode; Rancode2 & anemia) with an AIC value of 179.4; 179; 178.8 & 179.3 compared to the exponential distribution with an AIC value 267.4; 267.1; 267.2 & 267.4 respectively.

Table 3.5: The survival distribution and equality of survival distributions.

	Log-likelihood	**AIC**	**p-value**	**Log-likelihood**	**AIC**	**p-value**	**Log-likelihood**	**AIC**	**p-value**	**Log-likelihood**	**AIC**	**p-value**	
Distribution	Gender			Rancode			Rancode2			Anemia			Rank
Student-t	-86.7	179.4	0.93	-86.4	178.8	0.44	-86.5	179	0.52	-86.6	179.3	0.69	1
Logistic	-87.7	181.4	0.86	-87.6	181.3	0.69	-88.7	183.4	0.67	-87.7	181.4	0.88	2
Loglogistic	-88.8	183.5	0.85	-88.7	183.4	0.69	-88.7	183.4	0.67	-88.8	183.5	0.88	3
Gaussian	-89	184	0.63	-89.1	184.3	0.95	-89.1	184.2	0.89	-89.1	184.3	0.99	4
Loggaussian	-90.8	187.3	0.61	-90.8	187.6	0.93	-90.8	187.6	0.9	-90.8	187.6	0.98	5
Rayleigh	-114.3	232.5	0.79	-114.1	232.2	0.52	-114.1	232.2	0.56	-114.2	232.5	0.72	6
Exponential	-131.7	267.4	0.75	-131.6	267.1	0.53	-131.6	267.2	0.56	-131.7	267.4	0.7	7

4. DISCUSSION

From the results given by log- rank test in table 3.2 and KM diagrams in figure3A-D, we can conclude that gender does not matter when it comes to recurrence of malaria, neither does treatment- rancodes. Though anemia is not statistically significant as the results shows (p- value = 0.816) but from the curve, only those who were anemic had recurrence. This confirms the limitation of log- rank test and Gehan- Beslow statistics tests.

Based on the results of various parametric distributions that can fit the model and the covariates for this data of recurrent malaria, Student-*t-distribution* is the most appropriate of all the other parametric distributions tested, with AIC value of 177.4 for the model and also having the lowest value for the respective covariates. This is a consistent result and shows that neither exponential nor gamma nor Weibull distributions which are common distributions in most statistical survival analysis is appropriate for this data. A point worth noting is the fact that Weibull distribution does not converge in this data for the specific covariates but fits the overall model with AIC value of 180.7, third in rank after logistic distribution with AIC value of 179.4, this was corroborated by the analysis of some common distributions in survival analysis, having considered their density, f(t), survival function, S(t) and hazard functions h(t).

The Student-*t-distribution* is not commonly used in literature probably due to the mathematical complexities involved. However, *R* software works it out and gives the output result. A more complex model is usually a better predictive model, (O'Meara, 1968) because more parameters allow for flexibility in how the model fits the data. This is one of the reasons why Student-*t* distribution has come out to be the best among the other parametric distributions, given that AIC = -2L($\hat{\theta}$) + 2p; where L($\hat{\theta}$) is the log-likelihood function value and p the number of parameters in the distribution.

Student-*t-distribution* is rarely used by most statisticians but it is prudent to note here that Student- *t* tends to standard normal distribution when the degree of freedom get larger and that it is also bell- shaped like any normal distribution. In this study, the Student *t*-distribution appropriately fits the data since in recurrence of malaria and according to the study design, the tail gets longer with fewer patients remaining with recurrent malaria as the rest experience complete clearance with use of more effective antimalarials.

As some information may be lost due to censoring(Latimer, 2013; Su, 2016); to put it plainly, a censored observation contains only partial information about the random variable of interest(Wienke, 2003), it is therefore needful to use methods that are more effective than the common KM (nonparametric distributions) and use prognostic covariate information to recover some of the information lost due to censoring(Lee & Go, 1997). Whereas the nonparametric analysis of survival distribution takes care of the real data as it is without exaggeration, it is limited in that it is not an effective tool with regard to forecasting and policy decision making. Parametric distribution assumes a homogeneous scenario which may not be the case at all times. Parametric approach provides an additional advantage in identifying prognostic or risk factors. Parametric distributions

also have the advantage of taking care of the inherent measurement errors in longitudinal covariates, which tend to bias the association parameters(Ozg, 2015). Consequently, they minimize biased predictions due to over- or under estimations of study parameters. It is necessary to correctly choose the right distribution to ensure decision making arising from the resulting model is more robust.

4.1 CONCLUSION

We can conclusively say that though the KM method may have some limitations just like the Log- rank test, they are useful in eliciting basic hazard and survival distribution functions for longitudinal data. However, since it is now known that measurement error in the longitudinal covariate biases (Carroll et al.,2006) the association parameter between these data towards zero (Prentice,1982), there is need to improve analysis of longitudinal and survival data(Ozg, 2015). In this study we established that it is important to identify the right distributions in the analysis to ensure use of the right model that will improve the accuracy of conclusions about the performance of the survival distribution curves. This will make obtained results more robust and hence more useful for healthcare decision making. Additionally, the importance of both the parametric and nonparametric survival distributions cannot be underestimated. An advantage of using a parametric distribution is that we can reliably predict time to event well after the period during which events occurred for our observed data, it is also now possible to extract all the statistical information such as mean, quantile, variability in relation to survival times consistently under one parametric model and this opens up the prospect of developing more powerful statistical models and tests for censored survival data frequently used in engineering and medicine(Su, 2016).

5. RECOMMENDATIONS

From the foregoing, we recommend further studies on recurrent events with emphasis on the application of various probability distributions and the nonparametric distributions to varied data subjected to different study designs, for example, the Student *t*-distribution in survival analysis of longitudinal data, especially where potential heterogeneity exist. There is need for more consultative efforts by the epidemiologist, the biostatistician, clinicians, general medical researchers and health policy makers and basic public health worker be made aware of the survival analysis, interpretation and implementation in the health decision making. Since the data of study had a short follow up period and may not reveal much, we therefore recommend data with a longer follow up period and larger population of study be explored.

Acknowledgements

I wish to acknowledge and thank Professor Charles Obonyo who undertook the initial study of malaria in Siaya County and gave me this as secondary data for analysis of recurrence of malaria; all his associates and parties involved in the collection of the primary data.

6. REFERENCES

Box-Steffensmeier, J. M., & De Boef, S. (2006). Repeated events survival models: The conditional frailty model. *Statistics in Medicine*, *25*(20), 3518–3533. https://doi.org/10.1002/sim.2434

Butler, E. L. (2011). Estimating the Survival Distribution of Aluminum Processing Pots.

Checchi, F., Roddy, P., Kamara, S., Williams, A., Morineau, G., Wurie, A. R., … Pinoges, L. (2005). Evidence basis for antimalarial policy change in Sierra Leone : five in vivo efficacy studies of chloroquine , sulphadoxine – pyrimethamine and amodiaquine, *10*(2), 146–153.

Exercise 4 : Multivariable analysis in R part 2 : Cox proportional hazard model. (n.d.), 1–8.

Hagar, Y., & Dukic, V. (n.d.). Comparison of hazard rate estimation in R.

Hofert, M. (2013). On Sampling from the Multivariate t Distribution, *5*(December), 129–136.

Latimer, N. (2013). NICE DSU TECHNICAL SUPPORT DOCUMENT 14 : SURVIVAL ANALYSIS FOR ECONOMIC EVALUATIONS ALONGSIDE CLINICAL TRIALS - EXTRAPOLATION WITH PATIENT-LEVEL DATA, (0).

Lecture 15 Introduction to Survival Analysis. (2004).

Lee, E. T., & Go, O. T. (1997). Survival Analysis in Public Health Research. *Annu Rev Public Health Review of Public Health*, *18*(1), 105–34. https://doi.org/10.1146/annurev.publhealth.18.1.105

Obonyo, C. (2006). *Malaria, anaemia and antimalarial drug resistance in African children*. Retrieved from http://www.researchgate.net/publication/200103469_Malaria_Anaemia_and_Antimalaria_Drug_Resistance_in_African_Children/file/a453cdfd3390b3a388c520898af2573f.pdf

Ozg, A. (2015). Longitudinal and survival statistical methods with applications in renal medicine by ¨ ur Asar Declaration of Authorship, (August).

Su, S. (2016). Flexible modelling of survival curves for censored data. *Journal of Statistical Distributions and Applications*. https://doi.org/10.1186/s40488-016-0045-0

Wienke, A. (2003). Frailty Models. *Demographic Research*, *49*(0), 0–13. https://doi.org/10.4054/DemRes.2000.3.8