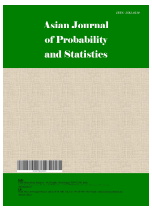


Asian Journal of Probability and Statistics

5(4): 1-18, 2019; Article no.AJPAS.52331
ISSN: 2582-0230

Survival Analysis of Cholera Patients a Parametric and Non-parametric Approach

Umar M. Hassan[1*] and A. A. Abiodun[2]

[1]Centre for Disaster Risk Management and Development Studies, University of Maiduguri, Maiduguri, Nigeria.
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2019/v5i430143
Editor(s):
(1) Dr. Jiteng Jia, School of Mathematics and Statistics, Xidian University, China.
Reviewers:
(1) Naresh Kumar, Mewar University, India.
(2) I. T. S. Piyatilake, University of Moratuwa, Sri Lanka.
Complete Peer review History: http://www.sdiarticle4.com/review-history/52331

Original Research Article

Received: 20 August 2019
Accepted: 27 October 2019
Published: 05November 2019

Abstract

Aims: The aim of this study is to investigate survival probability of cholera patients who were under follow-up and identify significant risk factors for mortality.
Methodology: In this research, we present the basic concepts, nonparametric methods (the Kaplan-Meier method and the log-rank test) and parametric method. Parametric AFT models (Exponential, Weibull, Lognormal and Log logistic) were compared using Akaike's Information Criterion (AIC).
Results: Recorded data of 513 patients were obtained from UNICEF Cholera Hospital for Internally Displaced Persons Camps within Maiduguri, Borno State. Non-Parametric and Parametric approach were used to estimate the survival probability of the patients and examine the association between the survival times with different risk factors. The analysis shows that some factors significantly contribute to longer survival time of cholera patients. These factors include being a female, age less than twenty, being vaccinated before the infection and mild degree of dehydration.
Conclusion: The vaccination, age, sex and degree of dehydration of a cholera patient affects its survival hence, much attention should be given to older patients, degree of dehydration and vaccine *(killed oral 01 with whole-cell with Bsubunit*) should be administered whenever there is outbreak. When carrying out survival analysis of this kind, a Weibull model is Recommended for used while if dealing with Accelerated Failure Time models.

**Corresponding author: E-mail: umarstat3@gmail.com;*

Keywords: *Cholera; parametric; non-parametric; event time ratio; exponential; weibull; lognormal; log logistic.*

1 Introduction

Cholera remains a global threat to public health and key indicator of lack of social development [1]. It is an acute enteric infection caused by ingestion of bacterium **vibrio cholera** present in faecal contaminated water and food. It has been classified as remerging global threat. The disease is primarily linked to insufficient access to safe clean water supplies, crowded living condition, poor hygiene and sanitation [1,2]. It has a more Severe impact in areas where basic environmental infrastructures is disrupted or have been destroyed [3].

Contaminated water is more common as usual transmission route in developing countries such as Nigeria [4].

Worldwide, about 1.4-4.3 million cases and 28,000-142,000 death per year are due to cholera infection. In Nigeria, cholera infection is endemic and outbreaks are common [5]. An upsurge of cholera cases was reported in September, 2013 by the Federal Ministry of Health and continued throughout December, 2013.

In 2016 a total of 6600 cholera cases, including 229 deaths (CFR 3.47%) were reported from 94 Local Government Areas in 20 states, Borno state inclusive. Borno State has been in the forefront of most recent cholera outbreak in Nigeria [6].

Ibrahim baffa sale, 2014 cholera line-list from the Kaduna State Disease Surveillance and Notification officer (DSNO). Described the outbreaks in time, place and person using Epi-info 7 and Health Mapper. Results: A total of 1468 case-patients and 54 deaths were recorded, giving a case fatality rate (CFR) of 3.68%. Female case-patients were 809(55.08%). The median age for case-patients was 15 years, with an age range of 0.04-90 years. Age specific case fatality rate (ASCFR) is highest among the > 60 years. The outbreak started from the first epidemic week of 2014 and lasted over 33 weeks. His analysis revealed a protracted cholera outbreak that gradually increases in magnitude throughout the first half of 2014 and spread within contiguous LGAs. He recommended the strengthening of the state's diseases surveillance system towards timely detection and early response to disease outbreaks in the future [7].

Adeneye in 2016, investigated the risk factors associated with cholera epidemic during the 2010 cholera outbreak in some States in Nigeria. Semi-structured questionnaires were administered to consented patients and/or their parents/guardians in Bauchi and Gombe States in North East Nigeria. Few (33.7%) respondents had access to safe and clean drinking water through the pipe-borne system compared to well (47.8%) and river (19.6%). Respondents' means of sewage disposal were: pit/latrine (77.2%); bush (15.2%); and water closet (4.3%). Only 34.8% knew water, food and poor sanitation as transmission routes for cholera. There was a significant gender difference in knowledge of lack of safe and clean drinking water and poor sanitation as contributing factors to cholera infection ($p<0.05$). Observation showed poor sanitation and food hygiene practices in the communities visited. The results provided insights for planning educational programmes through information, education and communication/behavioural change communication efforts to boost knowledge on cholera in the communities [8].

Shittu in, 2010 assess the epidemiological features of a GIS supported investigation of a cholera outbreak in Abeokuta. Abeokuta, the capital city of Ogun State, Southwestern Nigeria with an area of 1256 km2 and a population of 605, 451 people had an outbreak of cholera from 20th November, 2005 to 1st of January, 2006. The outbreak affected the Abeokuta North Local Government area where the municipal waterworks is located. Municipal water consumption was found to be associated with illness [McNemar's Chi square ($X2 = 20.5$; $p < 0.001$) and Odds ratio 10]. The epidemiological surveillance data showed a total of one hundred and fifteen cases and 11 deaths with case fatality rate of 9.6%. Male specific and female specific case fatality rates were 11.9 and 7.1%, respectively. The age group of 15 years and above accounted for 68.3% of the

cases and 90.9% of the deaths. Post epidemic environmental investigation showed progressive contamination along distribution points. Cholera is still a major cause of morbidity and mortality among youth and ageing population in Nigeria [9].

Variable Selection Method: The variables were selected based on the log-rank test. Level of each variable was compared by the log-rank, those with significant difference were considered for the analysis.

2 Methodology

2.1 Kaplan-meier estimator

The Kaplan-Meier estimator of the survival function or survival probability is defined as [10].

$$\hat{S}(t) = \Pi_{t_i \leq t}\left(\frac{n_i - d_i}{n_i}\right) \quad (1)$$

Where $n_i = n_{i-1}$ - d_{i-1} - c_{i-1} and has the convention that $\hat{S}$(t)= 1 if $t < t_1$
n_i Is number at risk up to time t_i, d_i is number of death at time t_i and c_i Number of censored observations at time t_i

2.2 Log-rank Test

The Log rank test, also referred to as the Mantel-Cox test, is the most widely used method of comparing two survival curves and can easily be extended to comparisons of three or more curves. 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. It is applicable to data where there is progressive censoring and gives equal weight to early and late failures. This statistic, like many other statistics uses the observed and expected cell counts over categories of outcomes where the categories for the log rank statistic are defined by each of the ordered failure times for the entire set of data being analysed. It is also a test used for comparing survival distributions for two or more groups and assumes that hazard functions for the two groups are parallel. When two groups are being compared, a statistic with 1 degree of freedom (known as 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. It can also be computed by dividing the square of the summed observed minus expected score for one of the groups by the variance of the summed observed minus expected score [11].

$$\text{Log}\, rank\ statistic = \frac{(O_i - E_i)^2}{Var(O_i - E_i)} \quad (2)$$

Where O_i is the observed counts and E_i is the expected counts.

2.3 Exponential AFT model

This distribution is characterized by a constant hazard rate λ , its only parameter. When λ is high it indicates high risk and short survival; when low, it indicates low risk and long survival [12].

$$S(t) = e^{-\lambda t}\ , t\ \geq 0 \quad (3)$$

Where t is specified survival time, λ is hazard rate can be reparametrized.

2.4 Weibull AFT model

This distribution is characterized by two parameters, γ and λ. The shape of the curve is determined by γ while λ determines its scale. When $\gamma = 1$, the hazard rate remains constant as time increases just like the exponential case. When $\gamma > 1$, the hazard rate increases while it decreases when $\gamma < 1$ as t increases. This

makes it possible for the Weibull distribution to be able to model survival distributions with increasing, decreasing or constant risk [13]. The following gives the survival of weibull distribution:

$$S(t) = e^{-\lambda t^{\gamma}} \tag{4}$$

Where t is specified survival time, λ is hazard rate can be reparametrized and γ is shape parameter.

2.5 Log- logistic AFT model

One limitation of the weibull hazard function is that it is a monotonic function of time. However, the hazard function can change direction in some situations. The log-logistic survival and hazard function is given by [14]:

$$S(t) = \frac{1}{1+\lambda t^{p}}, \quad t \geq 0, \quad \lambda > 0, \quad p > 0 \tag{5}$$

Where t is specified survival time, λ hazard rate can be reparametrized and p is shape parameter.

2.6 Lognormal AFT model

If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by [14]:

$$S_0(t) = 1 - \Phi\left(\frac{log\ t - i}{\sigma}\right) \tag{6}$$

Where i and σ are parameters, $\phi(x)$ is the probability density function and $\Phi(x)$ is the cumulative density function of the standard normal distribution. The survival function for ith individual is

$$S_i(t) = S_0\left(t/\dot{\eta}_i\right) \tag{7}$$

Where $\dot{\eta}_i$ acceleration factor can be reparametrized, t is specified survival time and S_0 baseline survival value.

2.7 Akaike's information criterion (AIC)

We can use statistical tests or statistical criteria to compare all these AFT models. Nested models can be compared using the likelihood ratio test. The exponential model, the Weibull model and log-normal model are nested within gamma model. For comparing models that are not nested, the Akaike information criterion (AIC) can be used instead, which is defined as

$$AIC = -2l + 2(p + k) \tag{8}$$

Where l is the log-likelihood, p is the number of covariates in the model and k is the number of model-specific ancillary parameters. The addition of $2(p+k)$ can be thought of as a penalty if non-predictive parameters are added to the model. Lower values of the AIC suggest a better model [7].

3 Data Analysis and Results

In this study, the subject of analysis was the data from 513 patients with cholera infection cases that had the event (death) and survive after treatment, which were collected from UNICEF cholera hospital for internally displaced persons (IDPs) camps within Maiduguri. From the data collected, it shows that patients were

diagnosed either having vaccinated or not vaccinated before the infection. Other variables that were used for analysis are Age, Gender, Marital status and vaccination status, degree of dehydration, Camps, and Outcome of the patients. Result of each method was performed by statistical package in R.

3.1 Descriptive and non-parametric analysis

It can be observe in the Fig. 1 above by the log-rank test, there is significant difference in the cumulative incidence of death among the cholera patients based on age group. The k-m curve also shows that patients in age group <1yr have better chance of surviving.

It can be observe in the Fig. 2 above by the log-rank test, there is significant difference in the cumulative incidence of death among the cholera patients based on sex. The k-m curve also shows that female have better chance of surviving.

It can be observe in the Fig. 3 above by the log-rank test, there is significant difference in the cumulative incidence of death among the cholera patients based on vaccination status. The k-m curve also shows that vaccinated patients have better chance of surviving.

It can be observe in the Fig. 4 above by the log-rank test, there is significant difference in the cumulative incidence of death among the cholera patients based on degree of dehydration. The k-m curve also shows that patients with mild (C) degree of dehydration have better chance of surviving.

Table 1. Summary results of cholera patient's death events by different demographic, health and risk behavior variables

	Status		Total 513(%)
Covariates	**Number censored (%)**	**Number of deaths (%)**	
Age groups			
<1yr	57(89.1%)	7(10.9%)	64(12.5%)
1-20yrs	134(87.6%)	19(12.4%)	153(29.8%)
21-40yrs	104(84.6%)	19(15.4%)	123(24.0%)
41-60yrs	71(71.0%)	29(29.0%)	100(19.5%)
>60yrs	48(65.8%)	25(34.2%)	73(14.2%)
Sex			
Male	200(77.5%)	58(22.5%)	258(50.3%)
Female	214(83.9%)	41(16.1%)	255(49.7%)
Vaccination status			
Vaccinated	318(96.1%)	13(3.9%)	331(64.5%)
Not vaccinated	96(52.7%)	86(47.3%)	182(35.5%)
Degree of dehydration			
A	10(12.2%)	72(87.8%)	82(16.0%)
B	157(86.3%)	25(13.7%)	182(35.5%)
C	247(99.2%)	2(0.8%)	249(48.5%)
Marital status			
Single	224(88.2%)	30(11.8%)	254(49.5%)
Married	161(79.7%)	41(20.3%)	202(39.4%)
Divorced	15(51.7%)	14(48.3%)	29(5.7%)
Widow	14(50.0%)	14(50.0%)	28(5.5%)
IDPs camp			
Muna	98(83.1%)	20(16.9%)	118(23.0%)
Dalori	116(79.5%)	30(20.5%)	146(28.5%)
Bakasi	81(78.6%)	22(21.4%)	103(20.1%)
Mule	119(81.5%)	27(18.5%)	146(28.5%)

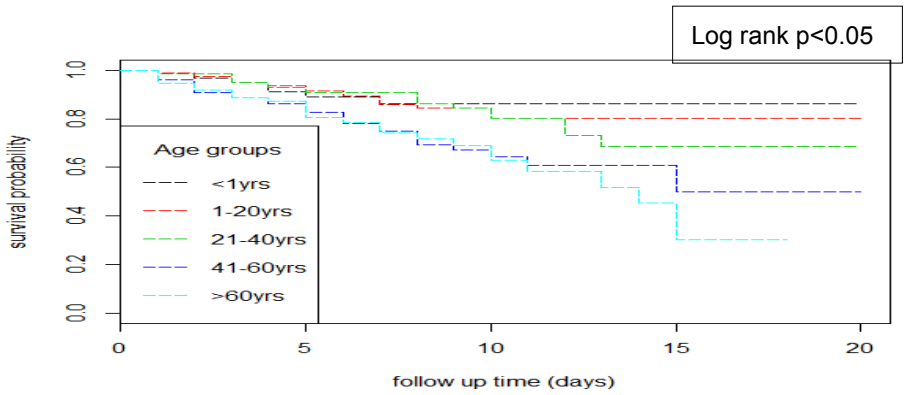


Fig. 1. Kaplan Meier curve based on age groups of the patients

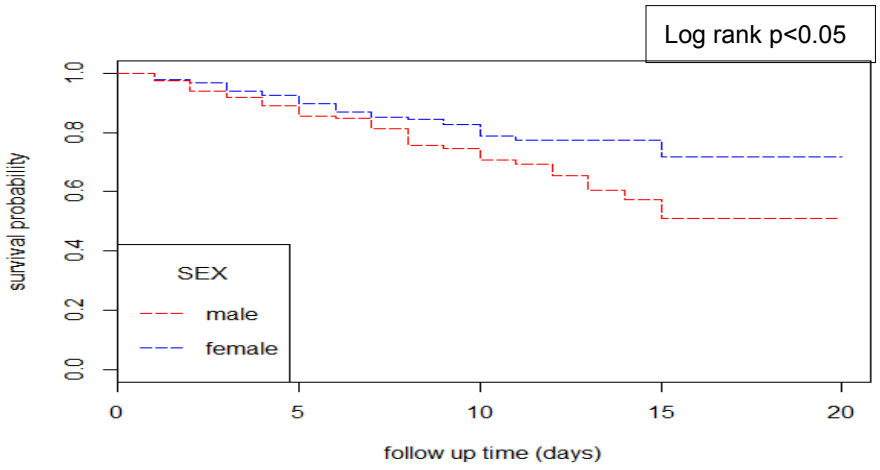


Fig. 2. Kaplan Meier curve based on sex the patients

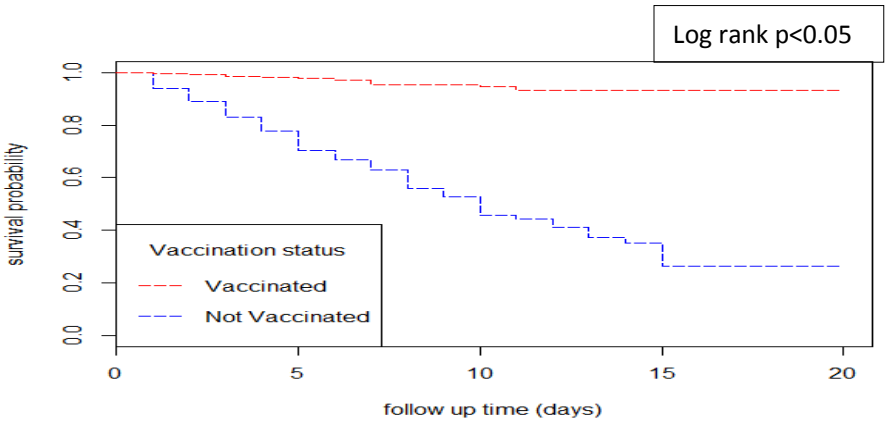


Fig. 3. Kaplan Meier curve based on vaccination status of the patients

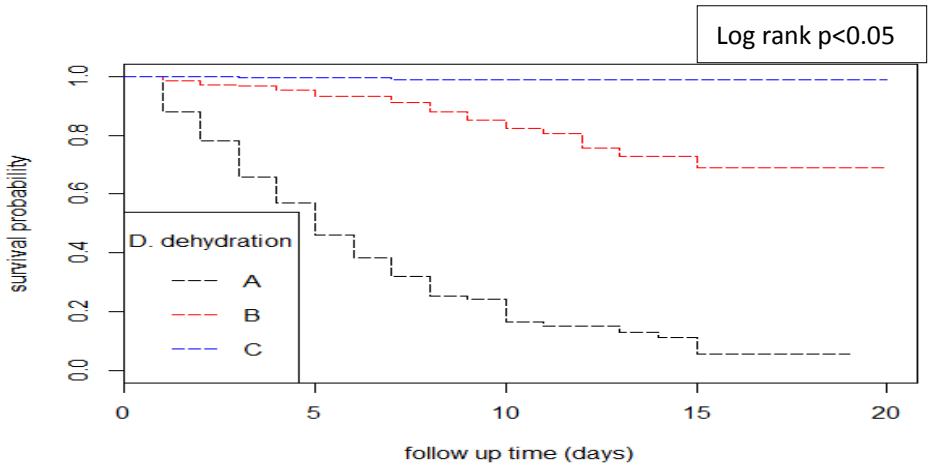


Fig. 4. Kaplan Meier curve based on degree of dehydration of the patients

Table 2. Results from AFT Models

Variable	Exponential		Weibull		Log-normal		Log-logistic	
	TR	**P**	**TR**	**P**	**TR**	**P**	**TR**	**P**
Intercept	**3.62**	**0.00**	**3.30**	**0.00**	**3.24**	**0.00**	**3.09**	**0.00**
Age groups								
21-40yrs	1		1		1		1	
<1yr	1.14	0.01	1.23	0.00	1.25	0.01	1.21	0.01
1-20yrs	1.42	0.04	1.33	0.02	1.32	0.03	1.40	0.03
41-60yrs	0.96	0.02	0.78	0.02	0.94	0.02	0.92	0.03
>60yrs	0.72	0.00	0.69	0.00	0.83	0.00	0.90	0.00
Sex	0.77	0.03	1.24	0.03	1.44	0.03	1.59	0.03
Vaccination status	1.63	0.00	1.75	0.00	1.92	0.00	1.83	0.00
Degree of dehydration								
Moderate (B)	1		1		1		1	
Severe (A)	1.64	0.01	0.85	0.00	1.63	0.01	1.76	0.01
Mild (C)	4.71	0.00	4.82	0.00	4.80	0.00	4.90	0.00
Scale	1		0.77		3.94		2.03	
Shape	1.00		1.29					
Log likelihood	-457.3		-453.2		-460.3		-458.3	

Table 3. The log-likelihoods and Akaike's Information Criterion (AIC) in the AFT models

Distribution	Log-likelihood	P	K	AIC
Exponential	-457.2884	8	1	932.5768
Weibull	-453.1534	8	2	926.3068
Lognormal	-460.3213	8	2	940.6426
Log-logistic	-458.5099	8	2	937.0198

3.2 Parametric analysis

We compared all these AFT models using statistical criteria (likelihood ratio test and AIC). The nested AFT models can be compared using the likelihood ratio (LR) test. The exponential model, the Weibull model and

the log-normal model are nested within the gamma model. However, the LR test is not valid for comparing models that are not nested. In this case, we use AIC to compare the models (Table 3) (The smaller AIC is the better). The weibull AFT model appears to be an appropriate AFT model according to AIC compared with other AFT models, although it is only slightly better than Exponential model. We also note that the log-normal model is poorer fits according to AIC.

Under the weibull AFT model (from Table 2), the estimated time ratio (TR) for patients in age group <1yr, 1-20yrs, 41-60yrs and >60yrs relative to patients in average age group (21-40yrs) is 1.23, 1.33, 0.78, 0.69 respectively. This indicates that the effect of <1yr, 1-20yrs prolongs the time to death, but the effect of 41-60yrs and >60yrs speeds up the time to death.

Being a Female significantly increase the survival time by approximately 25% (ETR =1.24) compared to male. Being vaccinated significantly increase the survival time by 75% (ETR = 1.75) compared to not vaccinated.

The patients with severe (A) degree of dehydration have shorter survival time than patients with moderate (B) degree of dehydration, but the patients with mild (C) dehydration have longer survival than patients with moderate (B) degree of dehydration.

4 Conclusion

Analysis was carried out to check whether the Age, Sex, Vaccination status and Degree of dehydration have a significant effect on the survival of cholera infected person.

Result shows difference in survival rate between the age groups, sex, vaccination status and degree of dehydration. Based on the log rank test carried out, patients of age (<1yr). Being a Female, being vaccinated before the infection and patients with mild degree of dehydration have better chance of surviving.

Four diverse models were applied to the dataset, such as Exponential AFT model, Weibull AFT model, Log-normal AFT model and Log-logistic. Among all the models, Weibull AFT model fits better and describes the data best.

Based on the weibull model Indicates that patients in age group <1yr, 1-20yrs have longer survival time relative to patients in age group (21-40yrs), but patients in age group 41-60yrs , >60yrs have shorter survival time relative to patients in age group (21-40yrs).

Being a Female significantly increase the survival time by approximately 25% compared to male. Being vaccinated significantly increase the survival time by 75% compared to not vaccinate. The patients with severe (A) dehydration have shorter survival time compared to patients with moderate (B) dehydration, but the patients with mild (C) dehydration have longer survival time compared to patients with moderate dehydration.

In conclusion, the Vaccination, age, sex and Degree of dehydration of a cholera patient affects its survival hence, much attention should be given to older patients, degree of dehydration and vaccine *(killed oral 01 with whole-cell with Bsubunit*) should be administered whenever there is outbreak.

When carrying out survival analysis of this kind, a Weibull model is Recommended for used while if dealing with Accelerated Failure Time models.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] World Health Organization Weekly epidemiological records. 2010; cholera. Geneva: World Health Organ. 2009;31:293-308.

[2] Sasaki. Spatial analysis of risk factor of cholera outbreak for 2003– 2004 in a peri-urban area of Lusaka, Zambia. Am. J. Trop. Med. Hyg. 2008;79(3):414-42.

[3] Cholera annual report. Weekly Epidemiological Record VW. Press GWHO. PubMed |Google Scholar; 2011.

[4] Sack Cholera. Lancet. 2004;363:897-898.

[5] National Population Commission Nigeria; 2016.
[Accessed 25 December]

[6] Hosmer DW, Lemeshow S, May S. Applied Survival analysis: Regression Modelling of Time to Event Data, Second Edition, John Wiley and Sons Inc., New York; 1998.

[7] Ibrahim baffa sale. Descriptive epidemiology of a cholera outbreak in Kaduna State, Northwest Nigeria; 2014.

[8] Adeneye. Risk factors associated with cholera outbreak in Bauchi and Gombe States in North East Nigeria; 2016.

[9] Shittu. Epidemiological features of a GIS-supported investigation of cholera outbreak in Abeokuta, Nigeria; 2010.

[10] Mantal Nathan. "Evaluation of survival data and two new rank order statistics arising in its consideration"; 1966.

[11] Lee J. Wang. Statistical methods for survival data analysis; 2003.

[12] Weibull, statistical distribution function of wide applicability; 1939.

[13] Collet D. Modelling Survival Data in Medical Research. Chapman and Hall, London; 2003.

[14] Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control*, 1974; 19:716–23.

APPENDIXS

VALIDITY OF THE RESULT

NON PARAMETRIC ANALYSIS

#KAPLAN MEIER SURVIVAL CURVE OF THE PATIENTS BASED ON AGE GROUPED

```
> fit4<-survfit(recsurv~colera$Agegroup)
> plot(fit4,
+       xlab = "follow up time (days)", ylab = "survival probability", col = c(1,2,3,4,5),lty=5)
> leg.txt<-c("<1yrs","1-20yrs","21-40yrs","41-60yrs",">60yrs")
> legend("bottomleft", leg.txt,col = c(1,2,3,4,5),lty=5, title = "Age groups")
> summary(fit4)
```

Call: survfit(formula = recsurv ~ colera$Agrp)

colera$Agegroup=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	64	1	0.984	0.0155	0.954	1.000
2	60	1	0.968	0.0223	0.925	1.000
3	53	1	0.950	0.0284	0.896	1.000
4	49	2	0.911	0.0382	0.839	0.989
5	42	1	0.889	0.0430	0.809	0.978
7	31	1	0.861	0.0503	0.767	0.965

colera$Agegroup=2

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	153	2	0.987	0.00918	0.969	1.000
2	140	2	0.973	0.01341	0.947	0.999
3	130	3	0.950	0.01833	0.915	0.987
4	116	2	0.934	0.02136	0.893	0.977
5	101	2	0.915	0.02462	0.868	0.965
6	84	2	0.894	0.02845	0.840	0.951
7	74	3	0.857	0.03413	0.793	0.927
8	60	1	0.843	0.03643	0.775	0.918
10	41	2	0.802	0.04478	0.719	0.895

colera$Agegroup=3

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	123	2	0.984	0.0114	0.962	1.000
3	110	4	0.948	0.0207	0.908	0.989
4	97	2	0.928	0.0245	0.882	0.978
5	89	2	0.908	0.0280	0.854	0.964
8	61	3	0.863	0.0366	0.794	0.938
9	49	1	0.845	0.0399	0.771	0.927
10	40	2	0.803	0.0478	0.715	0.902
12	22	2	0.730	0.0657	0.612	0.871
13	16	1	0.684	0.0758	0.551	0.850

colera$Agegroup=4

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	100	4	0.960	0.0196	0.922	0.999
2	92	5	0.908	0.0293	0.852	0.967
3	82	2	0.886	0.0325	0.824	0.952
4	77	2	0.863	0.0355	0.796	0.935
5	70	3	0.826	0.0399	0.751	0.908
6	56	3	0.781	0.0452	0.698	0.875
7	48	2	0.749	0.0488	0.659	0.851
8	40	3	0.693	0.0549	0.593	0.809
9	32	1	0.671	0.0573	0.568	0.793
10	24	1	0.643	0.0613	0.533	0.775
11	19	1	0.609	0.0668	0.491	0.755
15	11	2	0.499	0.0895	0.351	0.709

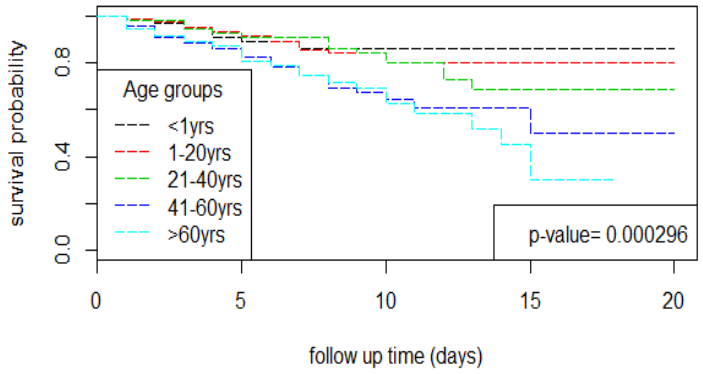
colera$Agegroup=5

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	73	4	0.945	0.0266	0.894	0.999
2	67	2	0.917	0.0325	0.856	0.983
3	63	2	0.888	0.0374	0.818	0.964
4	56	1	0.872	0.0399	0.797	0.954
5	52	4	0.805	0.0490	0.714	0.907
6	42	1	0.786	0.0514	0.691	0.893
7	37	2	0.743	0.0567	0.640	0.863
8	30	1	0.719	0.0600	0.610	0.846
9	26	1	0.691	0.0637	0.577	0.828
10	22	2	0.628	0.0718	0.502	0.786
11	14	1	0.583	0.0794	0.447	0.762
13	9	1	0.518	0.0934	0.364	0.738
14	8	1	0.454	0.1017	0.292	0.704
15	6	2	0.302	0.1105	0.148	0.619

> survdiff(recsurv~colera$Agrp,rho = 0)
Call:
survdiff(formula = recsurv ~ colera$Agrp, rho = 0)

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
colera$Agrp=1	64	7	12.3	2.30	2.70
colera$Agrp=2	153	19	28.6	3.22	4.65
colera$Agrp=3	123	19	25.0	1.42	1.95
colera$Agrp=4	100	29	19.0	5.24	6.67
colera$Agrp=5	73	25	14.1	8.40	10.06

Chisq= 21.1 on 4 degrees of freedom, p= 0.000296



#KAPLAN MEIER SURVIVAL CURVE BASED ON SEX OF THE PATIENTS

```
> fit1<-survfit(recsurv~colera$SEX)
> plot(fit1,
+      xlab = "follow up time (days)",ylab = "survival probability",col=c(4,2),lty = 5)
> leg.txt1<-c("male","female")
> legend("bottomleft",leg.txt1,col = c(2,4),lty = 5,title = "SEX")
> summary(fit1)
Call: survfit(formula = recsurv ~ colera$SEX)
```

colera$SEX=0

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	255	6	0.976	0.00949	0.958	0.995
2	233	2	0.968	0.01111	0.947	0.990
3	222	7	0.938	0.01564	0.907	0.969
4	198	3	0.923	0.01742	0.890	0.958
5	180	5	0.898	0.02037	0.859	0.939
6	158	5	0.869	0.02335	0.825	0.916
7	140	3	0.851	0.02521	0.803	0.902
8	120	1	0.844	0.02597	0.794	0.896
9	101	2	0.827	0.02802	0.774	0.884
10	84	4	0.788	0.03288	0.726	0.855
11	63	1	0.775	0.03465	0.710	0.846
15	27	2	0.718	0.05055	0.625	0.824

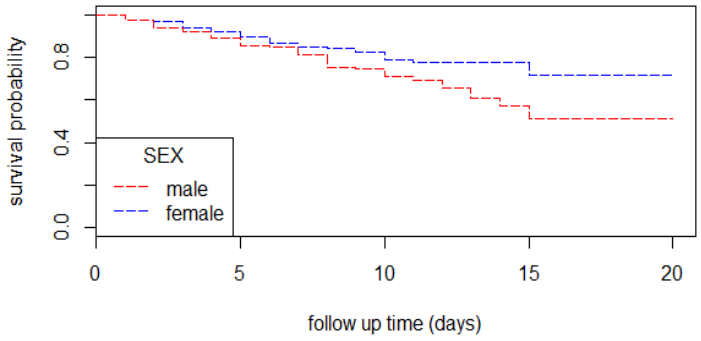
colera$SEX=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	258	7	0.973	0.0101	0.953	0.993
2	240	8	0.940	0.0149	0.912	0.970
3	216	5	0.919	0.0175	0.885	0.954
4	197	6	0.891	0.0203	0.852	0.931
5	174	7	0.855	0.0236	0.810	0.902
6	134	1	0.848	0.0243	0.802	0.897
7	120	5	0.813	0.0279	0.760	0.870
8	97	7	0.754	0.0336	0.691	0.823
9	79	1	0.745	0.0345	0.680	0.816
10	61	3	0.708	0.0387	0.636	0.788
11	45	1	0.693	0.0410	0.617	0.778
12	36	2	0.654	0.0469	0.568	0.753
13	27	2	0.606	0.0545	0.508	0.722
14	19	1	0.574	0.0602	0.467	0.705
15	18	2	0.510	0.0684	0.392	0.663

```
> survdiff(recsurv~colera$SEX,rho = 0)
Call:
survdiff(formula = recsurv ~ colera$SEX, rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
colera$SEX=0	255	41	52.4	2.48	5.44
colera$SEX=1	258	58	46.6	2.79	5.44

Chisq= 5.4 on 1 degrees of freedom, p= 0.0197



#KAPLAN MEIER SURVIVAL CURVE BASED ON DEGREE OF VACCINATION STATUS PATIENTS.

```
fit3<-survfit(recsurv~colera$VACCINATION)
> plot(fit3,
+      xlab = "follow up time (days)", ylab = "survival probability", col = c(4,2),lty = 5 )
> leg.txt<-c("Vaccinated","Not Vaccinated")
> legend("bottomleft", leg.txt,col = c(2,4),lty = 5, title = "Vaccination status")
> summary(fit3)
Call: survfit(formula = recsurv ~ colera$VACCINATION)
```

colera$VACCINATION=0

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	182	11	0.940	0.0177	0.906	0.975
2	165	9	0.888	0.0236	0.843	0.936
3	149	10	0.829	0.0285	0.775	0.887
4	131	8	0.778	0.0319	0.718	0.843
5	117	11	0.705	0.0357	0.638	0.779
6	95	5	0.668	0.0375	0.598	0.746
7	86	5	0.629	0.0391	0.557	0.711
8	71	8	0.558	0.0420	0.482	0.647
9	58	3	0.529	0.0430	0.451	0.621
10	44	6	0.457	0.0461	0.375	0.557
11	33	1	0.443	0.0468	0.360	0.545
12	28	2	0.412	0.0485	0.327	0.518
13	22	2	0.374	0.0508	0.287	0.488
14	18	1	0.353	0.0521	0.265	0.472
15	16	4	0.265	0.0547	0.177	0.397

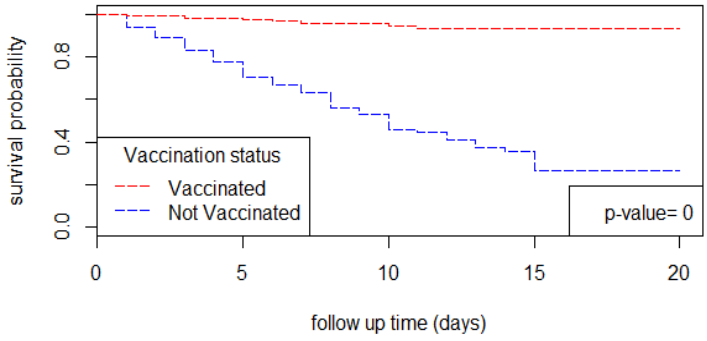
colera$VACCINATION=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	331	2	0.994	0.00426	0.986	1.000
2	308	1	0.991	0.00533	0.980	1.000
3	289	2	0.984	0.00717	0.970	0.998
4	264	1	0.980	0.00805	0.964	0.996
5	237	1	0.976	0.00902	0.958	0.994
6	197	1	0.971	0.01024	0.951	0.991
7	174	3	0.954	0.01390	0.927	0.982
10	101	1	0.945	0.01666	0.913	0.978
11	75	1	0.932	0.02066	0.893	0.974

> survdiff(recsurv~colera$VACCINATION,rho = 0)
Call:
survdiff(formula = recsurv ~ colera$VACCINATION, rho = 0)

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
colera$VACCINATION=0	182	86	33.1	84.6	131
colera$VACCINATION=1	331	13	65.9	42.5	131

Chisq= 131 on 1 degrees of freedom, p= 0



#KAPLAN MEIER SURVIVAL CURVE BASED ON DEGREE OF DEHYDRATION OF PATIENTS
fit2<-survfit(recsurv~colera$DEHYDRATION)
> plot(fit2,
+ xlab = "follow up time (days)", ylab = "survival probability", col = c(1,2,4),lty = 5)
> leg.txt<-c("A","B","C")
> legend("bottomleft", leg.txt,col = c(1,2,4),lty = 5, title = "D. dehydration")
> summary(fit2)
Call: survfit(formula = recsurv ~ colera$DEHYDRATION)

colera$DEHYDRATION=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	82	10	0.8780	0.0361	0.8100	0.952
2	72	8	0.7805	0.0457	0.6958	0.875
3	64	10	0.6585	0.0524	0.5635	0.770
4	53	7	0.5716	0.0548	0.4736	0.690
5	46	9	0.4597	0.0553	0.3631	0.582
6	36	6	0.3831	0.0542	0.2903	0.506

7	30	5	0.3193	0.0522	0.2318	0.440
8	25	5	0.2554	0.0489	0.1755	0.372
9	19	1	0.2420	0.0482	0.1638	0.357
10	16	5	0.1664	0.0434	0.0998	0.277
11	11	1	0.1512	0.0420	0.0877	0.261
13	8	1	0.1323	0.0408	0.0723	0.242
14	7	1	0.1134	0.0391	0.0577	0.223
15	6	3	0.0567	0.0303	0.0199	0.162

colera$DEHYDRATION=2

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	182	3	0.984	0.00944	0.965	1.000
2	168	2	0.972	0.01244	0.948	0.996
3	160	1	0.966	0.01376	0.939	0.993
4	151	2	0.953	0.01628	0.922	0.985
5	133	3	0.931	0.02010	0.893	0.972
7	98	2	0.912	0.02376	0.867	0.960
8	80	3	0.878	0.02998	0.821	0.939
9	67	2	0.852	0.03434	0.787	0.922
10	58	2	0.823	0.03893	0.750	0.903
11	44	1	0.804	0.04230	0.725	0.891
12	34	2	0.757	0.05136	0.662	0.864
13	26	1	0.728	0.05703	0.624	0.848
15	19	1	0.689	0.06564	0.572	0.831

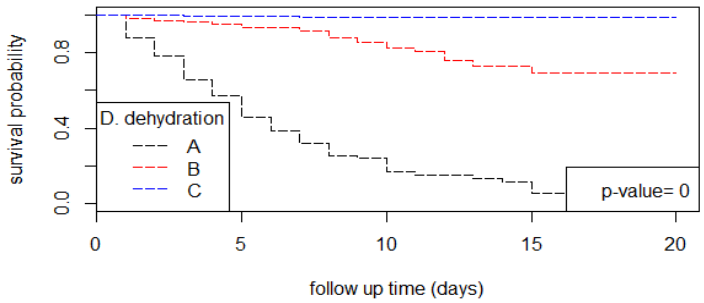
colera$DEHYDRATION=3

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
3	214	1	0.995	0.00466	0.986	1
7	132	1	0.988	0.00882	0.971	1

> survdiff(recsurv~colera$DEHYDRATION,rho = 0)
Call:
survdiff(formula = recsurv ~ colera$DEHYDRATION, rho = 0)

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
colera$DEHYDRATION=1	82	72	13.1	264.77	314.2
colera$DEHYDRATION=2	182	25	37.0	3.89	6.4
colera$DEHYDRATION=3	249	2	48.9	44.98	91.4

Chisq= 323 on 2 degrees of freedom, p= 0



PARAMETRIC ANALYSIS

#EXPONENTIAL MODEL USING EHA PACKAGE

colera<-read.csv("C:/Users/UMAR HASSAN/Documents/CHOLERA CODED.CSV")
> recsurv<-Surv(colera$Duration,colera$OUTCOME)
> library(eha)
> library(survival)
> AFTmodel1<-aftreg(Surv(colera$Duration,colera$OUTCOME)~Age + sex + dehydration + vaccination, data = colera,"exponential")
> summary(AFTmodel1)

Covariate	Mean	Coef	Exp(Coef)	se(Coef)	Wald p
Intercept	3.621				0.000
Age					
21-40yrs	0.124	0	1	(reference)	
<1yr	0.141	0.135	1.445	0.556	0.012
1-20yrs	0.289	0.356	1.248	0.549	0.043
41-60yrs	0.253	-0.038	0.963	0.477	0.021
>60yrs	0.193	-0.327	0.721	0.540	0.001
sex					
male	0.473	0	1	(reference)	
female	0.527	-0.258	0.772	0.209	0.032
dehydration					
B	0.133	0	1	(reference)	
A	0.375	0.495	1.642	0.251	0.011
C	0.892	1.549	4.713	0.730	0.000
Vaccination					
not vaccinated	0.333	0	1	(reference)	
vaccinated	0.667	0.552	1.750	0.330	0.000
log(scale)		1.000	2.718	0.411	0.007

Shape is fixed at 1

Events	99
Total time at risk	3693
Max. log. likelihood	-457.310
LR test statistic	246
Degrees of freedom	8
Overall p-value	0

#WEIBULL MODEL USING EHA PACKAGE
colera<-read.csv("C:/Users/UMAR HASSAN/Documents/CHOLERA CODED.CSV")
> recsurv<-Surv(colera$Duration,colera$OUTCOME)
> library(eha)
> library(survival)
> AFTmodel1<-weibreg(Surv(colera$Duration,colera$OUTCOME)~Age + sex + dehydration + vaccination, data = colera)
> summary(AFTmodel1)

Covariate	Mean	Coef	Exp(Coef)	se(Coef)	Wald p
Intercept	3.301				0.000
Age					
21-40yrs	0.124	0	1	(reference)	
<1yr	0.141	0.207	1.231	0.559	0.001
1-20yrs	0.289	0.285	1.330	0.655	0.020
41-60yrs	0.253	-0.041	0.960	0.461	0.020
>60yrs	0.193	-0.371	0.531	0.452	0.001
sex					
male	0.473	0	1	(reference)	
female	0.527	0.215	1.241	0.213	0.003
dehydration					
B	0.133	0	1	(reference)	
A	0.375	-0.162	0.851	0.252	0.000
C	0.492	1.572	4.821	0.730	0.000
Vaccination					
not vaccinated	0.333	0	1	(reference)	
vaccinated	0.667	0.559	1.750	0.334	0.000
log(scale)		-0.113	0.893	0.292	0.000
log(shape)		0.111	1.117	0.080	0.000

Events 99
Total time at risk 3693
Max. log. likelihood -453.243
LR test statistic 256
Degrees of freedom 8
Overall p-value 0

#LOGLOGISTIC MODEL USING EHA PACKAGE

AFTmodel1<-aftreg(Surv(colera$Duration,colera$OUTCOME)~Age + sex + dehydration + vaccination, data = colera, dist="loglogistic")
> summary(AFTmodel1)

Covariate	W.mean	Coef	Time-Accn	se(Coef)	Wald p
Intercept	3.091				0.001
Age					
21-40yrs	0.124	0	1	(reference)	
<1yr	0.141	0.190	1.210	0.501	0.011
1-20yrs	0.289	0.336	1.400	0.593	0.030
41-60yrs	0.253	-0.083	0.920	0.297	0.031
>60yrs	0.193	-0.105	0.900	0.287	0.001
sex					
male	0.473	0	1	(reference)	
female	0.527	0.464	1.591	0.154	0.031
dehydration					
B	0.133	0	1	(reference)	
A	0.375	0.565	1.760	0.172	0.010
C	0.492	1.524	4.590	0.417	0.000
vaccination					
not vaccinated	0.333	0	1	(reference)	
vaccinated	0.667	0.604	1.831	0.203	0.000
Baseline parameters:					
log(scale)		0.306		1.358	0.000
log(shape)		0.670		0.084	0.000
Baseline life expectancy:					

Events 99
Total time at risk 3693

Max. log. likelihood -458.312
LR test statistic 255
Degrees of freedom 8
Overall p-value 0

#LOGNORMAL MODEL USING EHA PACKAGE

AFTmodel1<-aftreg(Surv(colera$Duration,colera$OUTCOME)~Age + sex + dehydrtion + vaccination, data = colera, dist="lognormal")
> summary(AFTmodel1)

Covariate	W.mean	Coef	Time-Accn	se(Coef)	Wald p
Intercept	3.240				0.000
Age					
21-40yrs	0.124	0	1	(reference)	
<1yr	0.141	0.224	1.251	0.524	0.010
1-20yrs	0.289	0.207	1.230	0.515	0.031
41-60yrs	0.253	-0.061	0.940	0.322	0.021
>60yrs	0.193	-0.186	0.831	0.344	0.001
sex					
male	0.473	0	1	(reference)	
female	0.527	0.365	1.440	0.157	0.031
dehydration					
B	0.133	0	1	(reference)	
A	0.375	0.489	1.631	0.180	0.011
C	0.492	1.569	4.800	0.320	0.000
vaccination					
not vaccinated	0.333	0	1	(reference)	
vaccinated	0.667	0.652	1.921	0.197	0.000
Baseline parameters:					
log(scale)		1.145		0.298	0.000
log(shape)		0.045		0.074	0.540
Baseline life expectancy:					

Events 99
Total time at risk 3693
Max. log. likelihood -460.311
LR test statistic 249
Degrees of freedom 8
Overall p-value 0

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/52331