Modelling Internally Displaced Persons' (IDPs) Time to Resuming their Ancestral Homes after IDPs' Camps in Northern Uganda Using Parametric Methods

Joseph Okello Omwonylee[1], Abdou K. A. Diongue[2], Leo Odiwuor Odongo[3]

[1]Pan Africa University Institute of Basic Sciences, Technology and Innovation, P.O Box 62000-00200, Nairobi, Kenya

[2]Department of statistics at University of Gaston Berger, Senegal, Visiting lecturer to Pan Africa University Institute for Basic Sciences, Technology and Innovation, P.O Box 62000-00200, Nairobi, Kenya

[3]Kenyatta University Department of Mathematics, P.O Box 43844-00100, Nairobi, Kenya

Abstract: ***In this paper the time that the Internally Displaced Persons (IDPs) took to return from IDPs Camps to their ancestral homes in Northern Uganda was modeled using parametric methods. A retrospective dataset was collected from a cohort of 590 households belonging to seven different villages from Otuke districts that were displaced by Lord Resistance Army (LRA) war and were used in the study for a period of seven years. The raw dataset shows a total of 66 households that have not yet return to their ancestral homes. Stata inbuilt program and Easyfit 5.5 professional software were used to test the distribution (Exponential, Weibull and Log-logistic) fitness for the retrospective IDP dataset. The three distributions fit statistic of Kolmogorov-Smirnov, Chi-Squared and Anderson-Darling were used to test for the distribution fit. The Weibull distribution model was found to have a superior fit for the data than both exponential and log-logistic distribution model since it had a wider acceptance region from the test statistics. The study however recommended the widening of the scope for data collection in future studies for better statistical inference.***

Keywords: Survival techniques, parametric methods, Retrospective data, Internally Displaced Persons', Lord Resistance Army (LRA).

1. Introduction

The displacement of people into IDPs' Camps in Northern Uganda caused by Lord Resistance Army (LRA) war intensified in 2002. In 1987, LRA rebel group in Uganda was initiated by Joseph Kony believed to have been rooted in the rebellion against President Yoweri Museveni's National Resistance Movement (NRM) government but later transformed into a brutally violent war in which civilian turned to be the main victims, [3]. According to the same report, over 1.4 million people had been displaced and tens of thousands had been killed, raped or abducted by 2004. Over this period many people (12,000) have lost their lives, 20,000 abducted and 1.5 million lived in IDPs Camps [20]. According to [26] the LRA is led by Joseph Kony, who proclaims himself a spirit medium and apparently wishes to establish a state based on his unique interpretation of the Ten Commandments of God. According to report on Juba talks (2006-2008), peace talk were held that resulted into an agreement between Uganda government and the LRA rebels' ceasefire, the two parties signed a truce on 26 August 2006 while Uganda government began a process of creating 'satellite camps' to decongest the main IDPs camps and by mid-2007, thousands of IDPs had moved into the decongested camps. However, the populace remained cautious about the prospect of a peace deal with many refusing to return to their ancestral homes before definite end to the insurgency. This shows that the time the IDPs took to return to their ancestral homes varies and the technique of time-to-event analysis can be apply.

The 590 households of; Atat, Ogor, Omwonylee, Abur, Adodo, Otudu and Arom villages, in Otuke District were studied and a retrospective dataset for the period 2007 to 2013 was collected and modelled parametrically by Weibull, exponential and Log-logistics models.

2. Time-To-Event Analysis Techniques

According to [4], parametric, non-parametric and semi-parametric techniques are the three techniques used for analysing the time-to-event data, each with its own limitation but parametric approach is thought to yields better results provided the assumptions made in the analysis are correct. With Parametric models, the outcome is assumed to follow a certain known distribution. There are a number of texts that discuss comprehensively parametric time-to-event-models such as [7], [9], [10], [16], [19], [21], [22], [23] and [24]. For instance [19] suggest exponential, Weibull, lognormal and gamma distribution as the most commonly used parametric models in survival analysis.

Survival analysis is a phrase used to describe the analysis of data in the form of time from a well-defined time origin until the occurrence of the particular event of interest or the end point of the study [8]. It is a class of statistical techniques used for studying the occurrence and timing of events [2]. They were originally designed for the event of death occurrence and hence name survival analysis. The techniques is extremely useful for studying many different kinds of events in both the social and natural sciences, such as the onset of disease in Biostatistics, equipment failures in engineering, earthquakes, automobile accidents, stock market crashes, revolutions, job terminations, births, marriages, divorces, promotions in job places, retirements, Contracting Lung cancer due to smoking, arrests and many other time to event data. In Biostatistics, this techniques is often referred to as Clinical trials, in Engineering the term is referred to Reliability or failure time analysis, in econometric it is either duration analysis or transition analysis, and in Sociology it is often referred to as event history analysis. This is because the techniques have been adopted by researchers in several different fields. In this paper the time origin is when the Ugandan Government declared the villages safe in 2006 after signing of the truce and formation of satellite camps.

According to [25], Time-to-event analysis is frequently used with retrospective data in which people are asked to recall the dates when the events of interest happened to them (subject) which was the case in this study where subjects were asked to recall the year they return to their ancestral homes and the censored information was extracted from the record kept by the Local Council Chairperson hence a retrospective dataset was mined.

Several research have been conducted using the technique of (time-to-event) survival analysis such as [6], [11], [19], [23], [24], among others for many good case studies. Although much of the work in this paper pays much attention to internally displaced persons, the explored methods of parametric model are much more general. They can be applied to any study of time-to-event analysis, but in this work we consider the parametric models of Exponential, Weibull (two-parameters) and Log logistic (two-parameters).

3. Methodology

Sample of 590 households belonging to seven different villages in Otuke district were considered. The event of interest was returning to ancestral homes after displacement by LRA rebels. The censored subjects were those who might have died before returning to their ancestral homes during the seven years study period or a household who has not return to their ancestral homes between the study periods (2007-2013) yet the village was declare safe in the late 2006 after signing the truce between the Ugandan Government and LRA leaders. The accelerated failure time models (ATF) of Weibull, Exponential and log-logistic were considered. Exponential and Weibull model have been used in medical research in [23] and [24] to model survival data of CABG patients. On the other hands researcher such as in: [1], [8], [10], [12], [13], [15] [16], [17], and [18], provides literature on Weibull regression model, Exponential regression model and other parametric models with some including log-logistics distribution. In particular, [14] presented examples of the exponential, Weibull, and log logistic models and gave a brief description of other parametric approaches. The primary advantage of Weibull analysis is the ability to provide reasonably accurate failure analysis and failure forecasts with extremely small samples and providing a simple and useful graphical plot of the failure data [1]. Exponential is a special case of the Weibull distribution model. Graphical methods are used to test the assumption for the applicability of the three preferred distribution. These distributions were preferred because they are all accelerated failure time (AFT) models. For instance, log-logistics model can be implemented only for an Accelerated Failure time (AFT) form; Weibull and Exponential models can be implemented for both Proportional Hazard (PH) and AFT models. According to [1], normally distributed data makes a good Weibull plots but the reverse may always not be true.

4. Application

Detail application of Weibull and exponential distribution were presented in [12]. Exponential distribution is a special case of the Weibull distribution, when the shape parameter is equal 1. According to [14], the Weibull model has a key property that the log(−log) of S(t) is linear with the log of time allowing a graphical evaluation of the appropriateness of a Weibull model.

$$\hat{S}(t) = exp(-\lambda t^{\beta})\#$$
$$ln[-ln\,\hat{S}(t)] = ln(\lambda) + \beta\, ln(t) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.1)$$

Where β is the shape parameter (sometimes call the Weibull slope) and λ is the reciprocal of the scale parameter.

#

They further maintained that the log of failure odds for log logistics distribution is

$$\ln(\lambda t^{\beta}) = \ln(\lambda) + \beta ln(t) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.2)$$

Table 1: Displace Proportion of populace in a given time

Time (t)	*log-time* $[ln(t)]$	*Displaced portion* $[\hat{S}(t)]$	$ln\{-ln(\hat{S}(t))\}$	$\ln\left(\frac{1-\hat{S}(t)}{\hat{S}(t)}\right)$
1	0	0.7169	-1.1002	-0.9291
2	0.6931	0.4974	-0.3590	0.0104
3	1.0986	0.3385	0.0799	0.6700
4	1.3863	0.2757	0.2534	0.9659
5	1.6094	0.243	0.3469	1.1363
6	1.7918	0.1867	0.5178	1.4716
7	1.9459	0.1369	0.6874	1.8413

The natural logarithm of the time when the populace are displaced into the IDPs in camps, $[ln(t)]$ are taken to provide the horizontal axis of the test for the graphical assumption of the Weibull and Log-logistic regression appropriateness. The proportion displaced, $[\hat{S}(t)]$ of the IDPs from the Kaplan-Meier table is used to generate the column $[ln\{-ln(\hat{S}(t))\}]$ and $\ln\left(\frac{1-\hat{S}(t)}{\hat{S}(t)}\right)$ in Table.1: above which was later used for testing the Weibull and log-logistic regression appropriateness graphically as shown in Fig.1 and Fig.2 below.

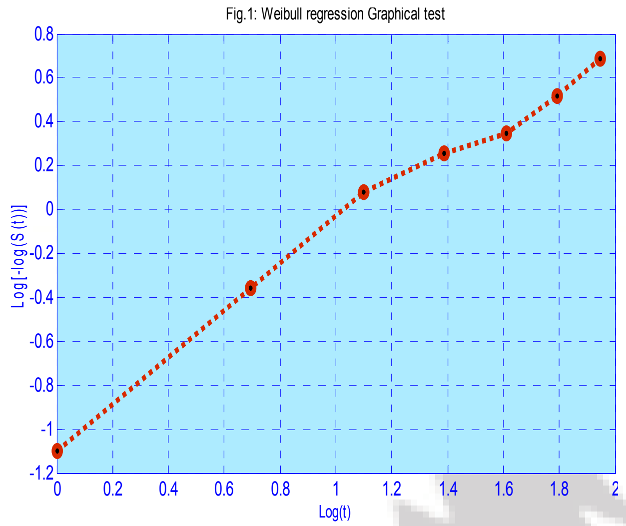

Figure 1: Shows the plot of $\ln[-\ln(\hat{S}(t))]$ against $\ln(t)$ and the result shows some straight line trend with a positive gradient evidencing that the data can be model with a Weibull regression model.

Table 2: Stata output for Weibull regression model

_t	*Coef.*	*Std. Err.*	*z*	*P>|z|*	*[95% Conf.*	*Interval]*
_cons	-1.81363	.08478	-21.39	0.00	.138098	.192538
/ln_β	.2941885	.035965	8.18	0.00	.223698	.364679
β	1.342037	.048267			1.25069	1.44005
1/β	.7451361	.026799			.694420	.799557

Weibull regression -- log relative-hazard form
No. of subjects = 590 Number of obs = 590
No. of failures = 494
Time at risk = 1857
LR chi2 (0) = -0.00
Log likelihood = -751.6265 Prob. > chi2 = .

Table.2: above is got by running the inbuilt Stata command **streg, dist(weibull) nohr.** Since 1.342037 (the shape parameter) is greater than 1 then the hazard is monotonically increasing with time, and therefore the rate of return to ancestral homes increases with time.

The obtained values of the shape and the scale parameters of a Weibull model are 1.342037 and 6.132654 respectively. Implying that the proportion of the displaced person at certain time is

$$\hat{S}(t) = \exp\left[-\left(\frac{t}{6.132654}\right)^{1.342037}\right] \quad \ldots\ldots\ldots\ldots 4.3$$

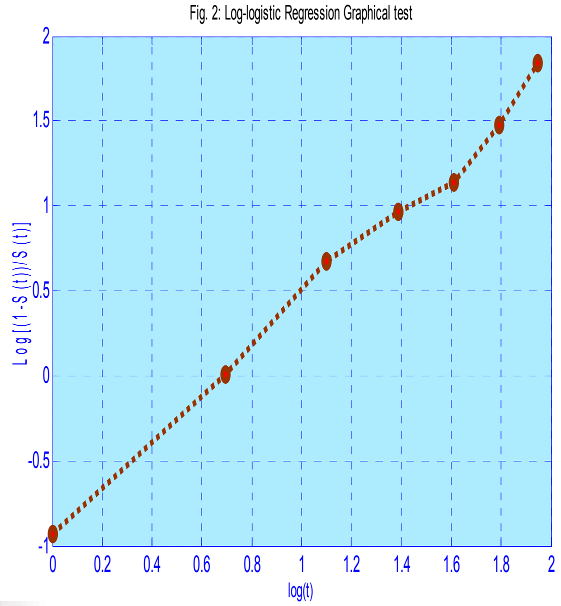

Figure.2. Shows that when $\log\left(\frac{1-\hat{S}(t)}{\hat{S}(t)}\right)$ is plotted against $\log(t)$ there is some trend of a straight line graph with a positive gradient showing that the time the IDPs take to return to their ancestral homes can be model by log-logistic Regression model as well.

Table 3: Stata output for Log-logistic regression model

_t	**Coef.**	**Std. Err.**	**Z**	**P>|z|**	**[95% Conf.**	**Interval]**
_cons	.9446	.0363	26.04	0.0	.8735	1.0157
/ln_gam	-.7020	.0366	-19.19	0.0	-.7736	-.6302
gamma	.4956	.0181			.5325	.5325

Log logistic regression -- accelerated failure-time form
No. of subjects = 590 Number of obs = 590
No. of failures = 494
Time at risk = 1857
Wald chi2 (0) = .
Log likelihood = -713.93399 Prob. > chi2 =.

Table.3: above is got by running the inbuilt Stata command **streg, dist (loglogistic) nolog.** Since gamma (0.4956235) is less than 1, then the hazard shape increases then decreases. This could justify the poor living condition experience in IDPs' camps which make people too eager to return to their ancestral homes. However those who could have had relatives and friends in the cities or those who were economically stable could return at will or when there is sure peace because they can afford rent in Town and afford other basic services. Since Stata provides estimates of the reciprocal of the slope parameter [14] then

$\beta = \frac{1}{gamma} = \frac{1}{0.4956235} = 2.01766$ is the slope and the scale parameter was estimated to be 2.9427 using Easyfit 5.5 professional software.

The log logistics distribution function is defined by

$$f(t;\beta,\alpha) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}}{\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]^{2}} \ldots\ldots\ldots\ldots 4.4.$$

For the positive shape, β and scale parameter, α as well as positive time, t with the proportion of the person displaced at a given time t defined by:

$$S(t;\beta,\alpha) = \frac{1}{\left[1+\left(\frac{t}{\alpha}\right)^{\beta}\right]} \ldots\ldots\ldots\ldots 4.5.$$

Table 4: Stata output for Exponential regression model

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
_con	-1.3242	.04499	-29.43	0.00	-1.4124	-1.236

Exponential regression -- log relative-hazard form
No. of subjects = 590 Number of obs = 590
No. of failures = 494
Time at risk = 1857
LR chi2 (0) = 0.00
Log likelihood = -780.81141 Prob. > chi2 = .

Table.4: above is got by running the inbuilt Stata command **streg, dist(exponential) nohr** on the time the displaced persons take to return to their ancestral homes after declaring the data set in the computer memory as a survival data using **stset** Stata command. The hazard rate was calculated as $Hazard\ rate = exp(-1.324182) = 0.266020 = \lambda$#
¹Ŧı ˝ˇ#ᵃ ˇ#proportional hazard model was used.
#
Since the hazard function is considered to be constant when a given dataset follows an exponential regression model, then h(t)=0.26602 has an exponential plot as shown in figure 3: below.

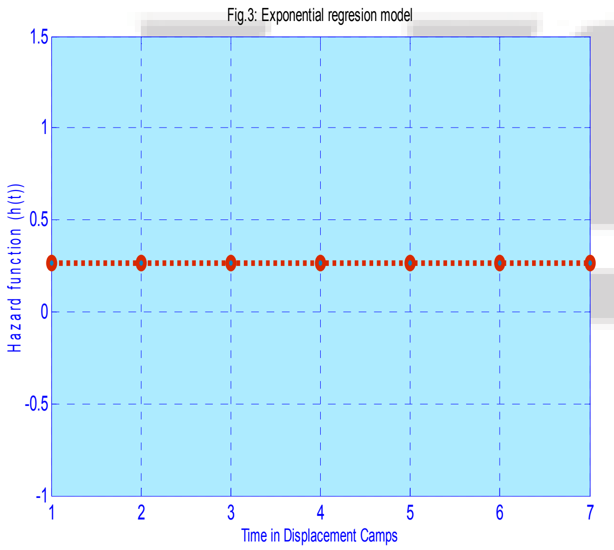

Figure 3: shows that the rate at which IDPs are returning to their ancestral home is constant with the hazard rate of 0.266020 which was got through the Stata software. The proportion of displaced Persons at a given time is then defined by:

$$S(t) = exp(-0.26602t) \ldots\ldots\ldots\ldots 4.6.$$

This proportion of displaced person when evaluated for the absolute error of 5×10^{-5} shows that no household will be displaced in 38 years' time i.e. by 2044 all natives would have returned to their ancestral homes.

4.1. The Model Results

Using the three parametric regression models (Weibull, Log logistic and Exponential) we obtain three different functions for the proportion of the displaced persons. These functions can be referred to as the survival functions if we were in Biostatistics research or reliability functions had we been dealing with equipment durability in engineering.

Using the proportion of the displaced persons function for the Weibull model from equation 4.3 and evaluating it at (four decimal places) as the absolute error, we project that no household will be displaced in 34 years' time from the start of the study i.e. by 2040 all natives would have returned to their ancestral homes.

Using the proportion of the displaced persons function for the Exponential model from equation 4.6 and evaluating it at (four decimal places) as the absolute error, we project that no household will be displaced in 38 years' time from the start of the study i.e. by 2044 all natives would have returned to their ancestral homes.

Using the proportion of the displaced persons function for the Log logistic model from equation 4.4 and evaluating it at (four decimal places) as the absolute error, it was not very clear when all natives would returned to their ancestral homes. Convergence to zero was when time is five hundred years showing a very weak convergence power.

Using our models we find out the proportion of the displaced persons as shown in table 5 below.

Table 5: Proportion of displaced person during the study period

Weibull regression model $\alpha = 6.1327$ $\beta = 1.3420$		Exponential regression model $\alpha = 0.2660$	Log-logistics regression model $\alpha = 2.0177$ $\beta = 2.9427$
Time (years)	Proportion displaced	Proportion displaced	Proportion displaced
0	1.0000	1.0000	1.0000
1	0.9160	0.7664	0.8982
2	0.8007	0.5874	0.6855
3	0.6818	0.4502	0.4903
4	0.5692	0.3450	0.3499
5	0.4672	0.2645	0.2555
6	0.3787	0.2027	0.1919
7	0.3029	0.1553	0.1482

4.2 Testing for the Graphical Fit Using Easyfit Software

Using Easyfit 5.5 professional and its inbuilt program of StatAssist 5.5, the null hypothesis that the IDPs dataset fits both Exponential and Log-logistic distribution are both rejected at the level of significances of 20%, 10%, 5%, 2% and 1% for all the three statistical test of Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared. On the other hand the null hypothesis for the fit of the IDPs data to two parameter Weibull distribution in not rejected for all the three statistical test of Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared at the level of significances of 20%, 10%, 5%, 2% and 1% . The IDPs dataset is therefore better modelled with Weibull distribution function. Table.6: Below shows the result from the Easyfit output.

Table 6: Easyfit output for Exponential, Weibull and Log logistic model

Exponential					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	590 0.3108 0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.04417	0.0504	0.0559	0.0625	0.06706
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size Statistic	590 103.19				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom Statistic P-Value	8 610.63 0				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	11.03	13.362	15.507	18.168	20.09
Reject?	Yes	Yes	Yes	Yes	Yes
Weibull					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	590 0.03328 0.51968				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.0441	0.0503	0.0559	0.0625	0.0671
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic	590 0.8705				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value	9 13.793 0.12988				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	12.242	14.684	16.919	19.679	21.666
Reject?	Yes	No	No	No	No
Log-Logistic					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	590 0.09289 6.9787E-5				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.0442	0.0504	0.0559	0.0625	0.06706
Reject?	Yes	Yes	Yes	Yes	Yes
Anderson-Darling					
Sample Size Statistic	590 5.9486				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	Yes	Yes	Yes
Chi-Squared					
Deg. of freedom Statistic P-Value	9 42.165 3.0655E-6				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	12.242	14.684	16.919	19.679	21.666
Reject?	Yes	Yes	Yes	Yes	Yes

5. Conclusion and Recommendation

Since the Weibull shape parameter is not equal to one the IDPs dataset had better be modelled by Weibull regression and not exponential regression. The graphical test for the appropriateness of the three distributions looks satisfied proposing that the dataset fits the distribution. The Weibull distribution has a superior fits as opposed to Exponential and Log-logistic distributions. The Easyfit 5.5 professional software test support the graphical test but with the Weibull distribution having a wider acceptance region than both the exponential and the log-logistic distribution. This implies that the displaced persons' return time to their ancestral homes can be model better by Weibull distribution model. The Weibull model shows that by 2040 all displaced persons would have returned to their ancestral homes and the Exponential distribution model shows that they would have returned to their ancestral homes by 2044. However, the survival function of the log logistic model takes unnecessarily long convergence time to converge to zero.

There is however need for collaborative research involving the social scientist and psychologist with the objective of investigating the possible causes of delay to resume the ancestral homes and such study need to be funded. The study also needs to be expanded to a wider sample to ensure good inference.

References

[1] Abernathy, R. B. (1998). The New Weibull Handbook. 3rd ed. SAE Publications, Warren dale. PA.
[2] Allison, P.D. (1995). Survival Analysis using SAS; A practical Guide Cary, NC: SAS Institute.
[3] Behind the Violence: Causes, Consequences and the Search for Solutions to the War in Northern Uganda, Refugee Law Project of Makerere University, Uganda, February 2004.
[4] Buis, M.L., (2006), An introduction to Survival Analysis, Department of Social research Methodology, Vrije Universiteit Amsterdam.
[5] Ceasefire First on Kony Agenda, All Africa (The Monitor), 15 July 2006
[6] Collett D. (2003), Modelling survival data in medical research, second edition, Chapman and Hall/CRC
[7] Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data. Chapman and Hall, New York
[8] Crow, L.H. (1982), "Confidence Interval Procedures for the Weibull Process With Applications to Reliability Growth," Technometrics, 24(1):67-72.
[9] Crowder, M.J., Kimber, A.C., Smith, R.L. and Sweeting, T.J. (1991). Statistical Analysis of Reliability Data. Chapman Hall, London, U.K
[10] Gross A.J. and Clark V.A. (1975). Survival Distribution: Reliability Applications in the Biomedical Sciences Wiley
[11] Ibrahim J.G., Chen M.H., and Sinha D.(2005). Bayesian Survival Analysis. Springer
[12] Khan K.H, Saleem M and Mahmud. Z. (2011). Survival Proportions of CABG Patients: A New Approch. Volume 3, Number 3.
[13] Klein.P.J and Moeschberger.L.M (1997, 2003). Survival Analysis Techniques for Censored and Truncated Data.
[14] Kleinbaum D. and Klein M, (2005) Logistic Regression: A Self-Learning Text, Second Edition, Springer-Verlag Publishers, New York, Chapters 4–7, 11
[15] Kleinbaum, D.G. and Klein, M. (2005). Survival analysis: a self-learning text.
[16] Lawless Jerald F. Lawless (1982). Statistical Models and Methods for Lifetime Data, John Wiley and Sons, Inc New York.
[17] Lawless, J.F. (2003), Statistical models and methods for lifetime data, second edition, Wiley-Inter-Science, A John Wiley & Sons, Inc., Publication, Hoboken, New Jersey
[18] Lawrence M. Leemis (1995). Reliability Probabilistic Model and Statistical Methods
[19] Lee E.T and Wang J.W. (2003), Statistical methods for survival data analysis, 3rd edition, John Wiley & sons, Inc., Hoboken, New Jersey.
[20] Nampindo S., Phillipps G.P. and Plumptre A. (2005), The Impact of Conflict in Northern Uganda on the Environment and Natural Resource Management. USAID
[21] Nelson, W. (1982). Applied Life Data Analysis. New York: Wiley
[22] Nelson, W. (1990). Accelerated testing, John Wiley and Sons, New York
[23] Saleem M., Mahmud Z., and Khan K. H. (2012) Survival Analysis of CABG Patients by Parametric Estimations In Modifiable Risk Factors - Hypertension and Diabetes. American Journal of Mathematics and Statistics, 2012; 2(5): 120-128
[24] Saleem, M, Mahmud, Z and Khan, K.H. (2012). Maximum Likelihood Estimation of CABG Patients By Parametric Models Based on Incomplete and Complete Population. American Journal of Statistics and Mathematics (JOSAM): Bioinfo Publications.
[25] Selvin, S. (2008). Survival analysis for epidemiologic and medical research. A practical guide, Cambridge University Press, New York, USA.
[26] Ward, K. (2001). The Armies of the Lord: Christianity, Rebels and the State in Northern Uganda, 1986-1999, Journal of Religion in Africa 31 (2)

Author Profile

Joseph Okello Omwonylee received B.Sc. (Mathematics major) from Busitema University in 2010. He was later

retained as a teaching assistant and has been a staff since then. He is currently pursuing M.Sc. (Statistics) at Pan African University Institute for Basic Science, Technology and Innovation (PAUISTI) hosted by Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya.

Professore Abdou Ka Diongue is a professor at Department of statistics, faculty of applied science and technology of the

University Of Gaston Berger Senegal. He has published more than twenty papers and attended many research conferences. He is currently a visiting lecturer to Pan Africa University Institute for Basic Sciences, Technology and Innovation (PAUISTI) hosted by the Jomo Kenyatta University of Agriculture and Technology-Nairobi, Kenya where he handles the courses of Statistical inference and Non-Parametric methods.

Prof Leo Odiwor Odongo is a professor of statistics. He received PhD (Statistics) and M.Sc.

(Statistics) from University of Manitoba, Canada in 1993 and 1990 respectively, M.Sc, (Mathematical statistics) and B.Ed.(Sc.) respectively, M.Sc, (Mathematical statistics) and B.Ed.(Sc.)(Double Mathematics) from University of Nairobi in 1983 and 1980 respectively. He has published over ten papers, written two books and presented over 20 conference papers. He is currently a professor in the department of mathematics of Kenyatta University.