Application of Survival Analysis on Antenatal Appointment

Akpan N. P.[1,*], Bassey N. A.[2]

[1]Department of Mathematics and Statistics, University of Port Harcourt, Nigeria
[2]Department of Statistics, Akwa Ibom State college of Art and Science, Nung Ukim, Nigeria

Abstract One of the reasons pregnant women has complications during pregnancy or child delivery is negligence of antenatal appointments and inability to adhere to medical advice. Due to the inconsistencies on the antenatal appointments, data generated from it has not been frequently used for statistical analysis. In this study, the application of survival analysis on antenatal appointment data was a priority. The survivorship function of the data was estimated using Kaplan-Meier method. The results showed that pregnant women attend the antenatal care for an average period of six (6) months. Cumulative hazard curves were plotted and they revealed that the antenatal appointment data can best be analyzed using the Weibull survival function and Weibull hazard function with parameters $\alpha = 2.89$ and $\lambda = 0.004$.

Keywords Survival analysis, Antenatal appointment, Kaplan-Meier method

1. Introduction

Survival analysis is a specialized field of mathematical statistics developed to study a special type of random variable of positive values with censored observations of which survival time events are the most common (Ma and Krings 2008). The main variable of interest is time – to – event. Examples of time – to- event random variables are incubation time of disease such as AIDS, failure time of certain machine components, times to death of patients with certain disease, the length of time of economic recessions and the occurrence of the next traffic accident. Other kinds of events which could be studied by survival analysis are divorce, criminal recidivism, child bearing, unemployment and graduation from school (John Fox, 2014).

Initially, survival analysis was useful only in the field of biostatistics and was mainly applied in medical researches for better analysis of life table data. It was also used to study death as an event specific to medical studies and demographic studies (Narendranathan et. al. 1993). As from 1970 till date, the statistical technique have been frequently applied in several fields of human endeavour such as engineering, economics, social sciences, medicine and public health just to mention but few. Its name varies from discipline to discipline, in engineering, it is called failure time analysis, in sociology it is referred to as history analysis. It is called survival analysis in biostatistics and medicine.

The advantage of survival analysis is its ability to handle censoring in an effective manner (Ma and Krings, 2008) although observation of censoring or incomplete information are not necessary for the application of survival analysis.

2. Materials and Methods

The data used in this study were obtained from the record unit of Poly clinic, Ini Local Government Area of Akwa Ibom State, Nigeria. The data set consisted of 1311 pregnant women who visited the clinic for booking and antenatal appointment between the year 2009 to 2013.

There are several models used in survival analysis. Naturally any distribution of non-negative random variables could be used to describe durations. In this study, we are interested in parametric models such as the Exponential models, Weibull model and Gompertz models. This is because they have closed forms expressions for survival and hazard functions.

Survival function s(t):

This represents the probability that an individual survives from the time origin to some time beyond t. It is given by

$$S(t) = p(T>t)$$
$$= \int_t^\infty f(t)dt$$
$$= 1 - f(t) \tag{1}$$

Hazard function λ(t):

This is the probability that an individual stops attending antenatal care at time t, given that she had survived up to that time. It is linked to the probability density function and survival function as follows:

$$\lambda(t) = \lim_{\Delta t \to 0} \Pr(t \le T \le t + \Delta t / T \ge t) / \Delta t$$

$$= \frac{f(t)}{S(t)}$$

$$= -\frac{d[Log\ s(t)]}{dt} \qquad (2)$$

Kaplan – Meier Estimator:

Suppose the events of interest occur at time

$t_1 < t_2 < ---- < t_i < --------- < t_n$, then the Kaplan Meier estimate of the survival function is given by;

$$\hat{S}(t) = \prod_{ti<t} \left(1 - \frac{di}{ni}\right)$$

$$= \prod_{ti<t} \{ni - di\}/ni \qquad (3)$$

where n_i is the number of pregnant women who continue their antenatal care at time t_i.

d_i is the number of those who stopped their antenatal care at time t_i.

Exponential Distribution

The exponential distribution model $(t \sim \exp(\lambda))$ is the simplest model and assumed a constant risk over time. The probability to event of interest within a particular time interval depends only on the length but does not depend on the location of the interval.

$$\text{ie } p(t \le T < t + \Delta / T \ge t) = p(T < \Delta)$$

The parameter λ attains only positive values, and the following formula are associated with the distribution.

Probability density function: $f(t) = \lambda e^{-\lambda t}$ (4)

Survival function: $s(t) = e^{-\lambda t}$ (5)

Hazard function: $\lambda(t) = \lambda$ (6)

Cumulative hazard function: $\Lambda(t) = \lambda t$ (7)

Weibull Distribution:

The Weibull model is a generalization of the exponential model with parameters α and λ which are called shape and scale parameters respectively. The parameters determine the shape and scale of the hazard function. The convenience of the Weibull model for empirical work arises from the flexibility and simplicity of the hazard and survival function. Its functions are as follows:

Probability density function: $f(t) = \lambda \alpha\ t^{\alpha-1}\ e^{-\lambda t^{\alpha}}$ (8)

Survival function: $s(t) = e^{-\lambda t^{\alpha}}$ (9)

Hazard function: $\lambda(t) = \lambda \alpha\ t^{\alpha-1}$ (10)

Cumulative hazard function: $\Lambda(t) = \lambda t^{\alpha}$ (11)

Gompartz Distribution:

A random variable follows a Gompertz distribution with parameters $a>0$ and $b>0$ $(t \sim \text{Gompertz } (a,b))$ if the following relations hold:

Probability density function: $f(t) = ae^{bt}\ e^{-a/b(e^{bt}-1)}$ (12)

Survival function: $S(t) = e^{-a/b(e^{bt}-1)}$ (13)

Hazard function: $\lambda(t) = ae^{bt}$ (14)

Cumulative hazard function: $\Lambda(t) = a/b(e^{bt}-1)$ (15)

Plots for Parametric Models:

In order to investigate the appropriate parametric model that can be used for the analysis of the antenatal care data, we used a straight line fit. We used the cumulative hazard $\Lambda(t)$ which is a better determinant of the appropriate survival distribution because its form is more distinct than that of density function f(t) or the survival function s(t), (Gross and Clark, 1975).

Bradburn et.al (2003) stated that;

i.   The plots of log $\Lambda(t)$ against t is linear for exponential distribution.

ii. The plots of log $\Lambda(t)$ against log t is linear for Weibull distribution.

iii. The plots of log $\Lambda(t)$ against t is linear for Gompertz distribution.

Therefore, we plotted the graphs in order to obtain the suitable model and the survival function was used to estimate the cumulative hazard function thus;

$$\Lambda(t) = -\log_e \hat{S}(t) \qquad (16)$$

In plotting log $\Lambda(t)$ against log t, if the line is approximately straight then its slope gives an estimate of the shape parameter (α) while the exponential of the intercept gives an estimate of the scale parameter (λ) of the Weibull distribution. Also, if the slope of the line is approximately unity (1), then the exponential model is considered appropriate. (Collett D, 1994).

In order to estimate the median survival time, we draw a horizontal line from the vertical axis to the curve at $\hat{S}(t) = 0.5$, the point of intersection with the curve determines the median.

The mean survival time based on the Kaplan – Meier product limit is obtained by;

$$\mu = 1.0t_1 + \hat{S}(t_1)\ (t_2-t_1)\ t - - - - + \hat{S}(t_{n-1})\ (t_n-t_{n-1}) \qquad (17)$$

3. Analysis/Discussion

ABBREVIATIONS:

Table 1. The Antenatal Data

t_i	n_i	d_i	Censored
0-1	1311	0	0
1-2	1311	12	5
2-3	1294	26	8
3-4	1260	59	31
4-5	1170	98	71
5-6	1001	223	128
6-7	650	263	132
7-8	355	101	102
8-9	152	67	63
9+	22	0	22

t_i -Survival period in months

n_i -Number of pregnant women who were able to continue their antenatal appointment for the period t_i.

d_i -Number of pregnant women who were unable to continue their appointment.

Censored - Number of pregnant women who delivered in the period t_i.

From equation (3), the unconditional probability that a pregnant woman will continue her antenatal appointment beyond any time t_i is estimated as follows:

$$\hat{S}(1) = \frac{1311 - 12}{1311} = 0.991$$

$$\hat{S}(2) = 0.991 \times \frac{1294 - 2}{1294} = 0.971$$

$$\hat{S}(3) = 0.971 \times \frac{1260 - 59}{1260} = 0.926$$

$$\hat{S}(4) = 0.926 \times \frac{1170 - 98}{1170} = 0.848$$

$$\hat{S}(5) = 0.848 \times \frac{1001 - 223}{1001} = 0.659$$

$$\hat{S}(6) = 0.659 \times \frac{650 - 163}{650} = 0.494$$

$$\hat{S}(7) = 0.494 \times \frac{335 - 101}{335} = 0.353$$

$$\hat{S}(8) = 0.353 \times \frac{152 - 67}{152} = 0.197$$

$$\hat{S}(9) = 0.197 \times 1 = 0.197$$

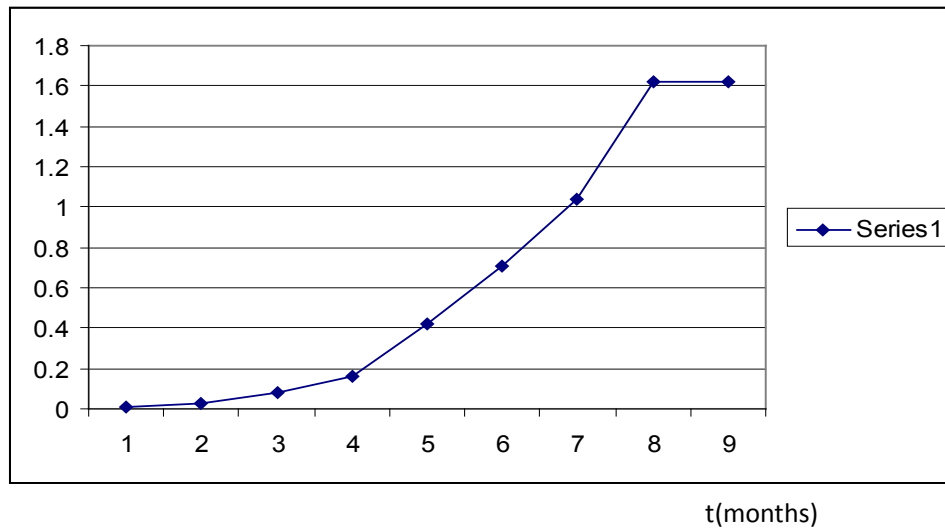Table 2 for plotted values.

$\Lambda(t)$

t(months)

Figure 1
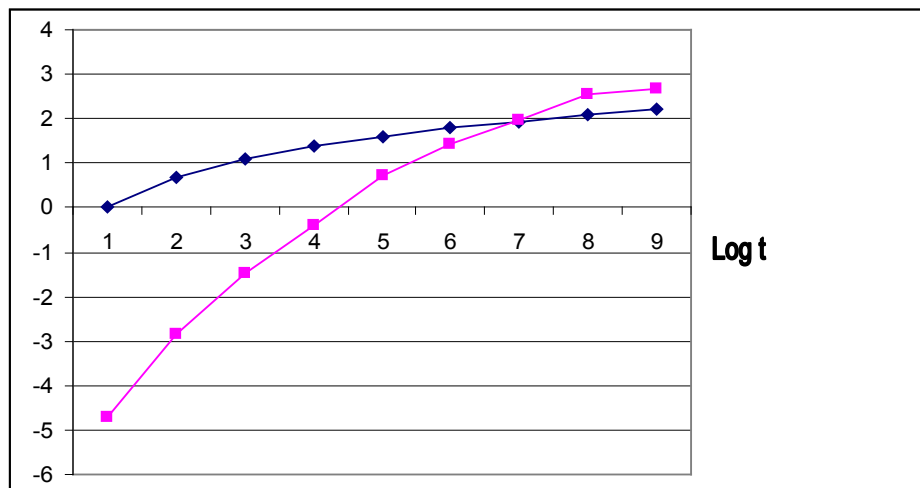
Plotting for Weibull

Log $\Lambda(t)$

Log t

Figure 2

Plotting for Gompartz

Log Λ(t)

(t)

Figure 3

Table 2. Plotting for Exponential Distribution

t	Logt	Ŝ(t)	Λ(t) = - logₑ Ŝ(t)	LogΛ(t)
0				
1	0	0.991	0.009	-4.711
2	0.693	0.971	0.029	-3.540
3	1.099	0.926	0.077	-2.564
4	1.386	0.848	0.165	-1.802
5	1.609	0.659	0.417	-0.875
6	1.792	0.494	0.705	-0.350
7	1.946	0.353	1.041	0.040
8	2.079	0.197	1.625	0.486
9	2.197	0.197	1.625	0.486

From the graphs, the plots of cumulative hazard ($\Lambda(t)$) against time (t) and the plots of logarithm of cumulative hazard, $\log\Lambda(t)$ against time (t) are not linear, but the plot of logarithm of cumulative hazard, $\log\Lambda(t)$ against logarithm of time, $\log(t)$ is linear which implies that the ante-natal data can be modeled by using Weilbull survival distribution with shape parameter ($\alpha = 2.89$) and scale parameter ($\lambda = 0.004$).

Therefore, the antenatal data can be modeled as:

(i) Pdf: $f(t) = 2.89\ \lambda t^{2.89-1}\ e^{-\lambda t^{\wedge}2.89}$

(ii) Survival function: $S(t) = e^{-\lambda t^{\wedge}2.89}$
(iii) Hazard function: $\lambda(t) = 2.89\ \lambda t^{2.89-1}$
(iv) Cum. Hazard function: $\Lambda(t) = \lambda t^{2.89}$
    Where $\lambda = 0.004$

The estimated mean survival time for the antenatal data is:

μ = 1.0(1) + (0.991) (2-1) + (0.971) (3-2)
    + (0.926)(4-3) + (0.848)(5-4) + (0.659)(6-5)
    + (0.494)(7-6) + (0.353)(8-7) + (0.197)(9-8)
    = 6.439
    = 6 months

Mean survival time for teenager:

μ = 1.0(1)+0.994+0.988+0.965+0.860+0.716+0.595
    +0.478+0.295
   = 6.891

Mean survival time for adults:

μ = 1.0 (1)+0.990+0.965+0.911+0.843+0.638+0.456
    +0.301+0.143
    = 6.247

The mean survival time for teenager is 6.891 (i.e approximately 7 months) while that of the adult is 6.247 (i.e approximately 6 months).
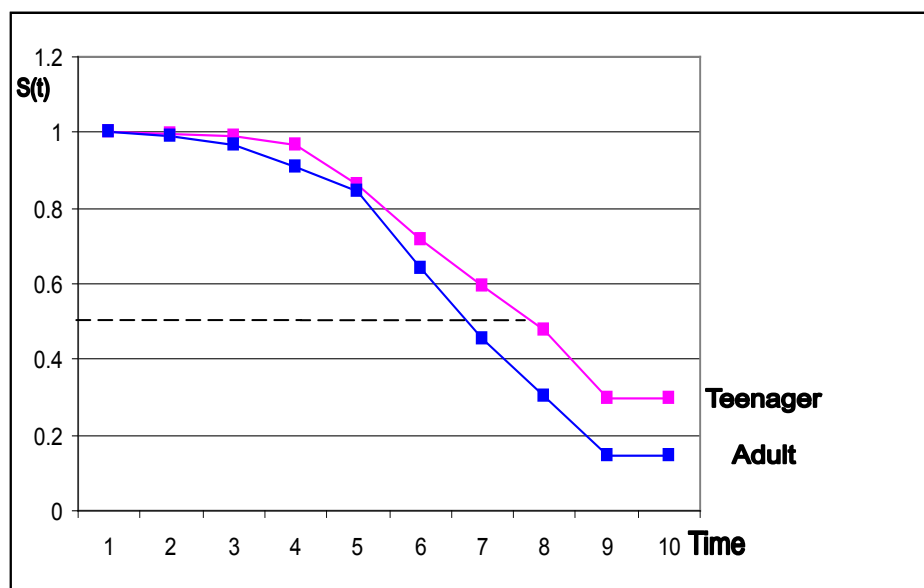
From figure 4, the median survival times for the teenager group and adult group are 7 months and 6 months respectively, while the median survival time for the entire pregnant women is 6 months as shown in figure 5.

Table 3.   Kaplan Meier Estimates for teenager (age<20 yrs)

Survival time(t)	No. at risk (n_i)	No unable to continue (d_i)	No. censored	K.M$\hat{S}$(t)
0	357	0	0	1
1	357	2	1	0.994
2	354	2	3	0.988
3	349	8	9	0.965
4	332	36	27	0.860
5	269	45	34	0.716
6	190	32	21	0.595
7	137	27	21	0478
8	89	34	39	0.295
9	16	0	16	0.295

Table 4.   Kaplan Meier Estimates for adults (Age≥20)

Survival time(t)	No. at risk (n_i)	No unable to continue (d_i)	No. censored	K.M$\hat{S}$(t)
0	954	0	0	1
1	954	10	4	0.990
2	940	24	5	0.965
3	911	51	22	0.911
4	838	62	44	0.843
5	732	178	94	0.638
6	460	131	111	0.456
7	218	74	81	0.301
8	63	33	24	0.143
9	6	0	6	0.143

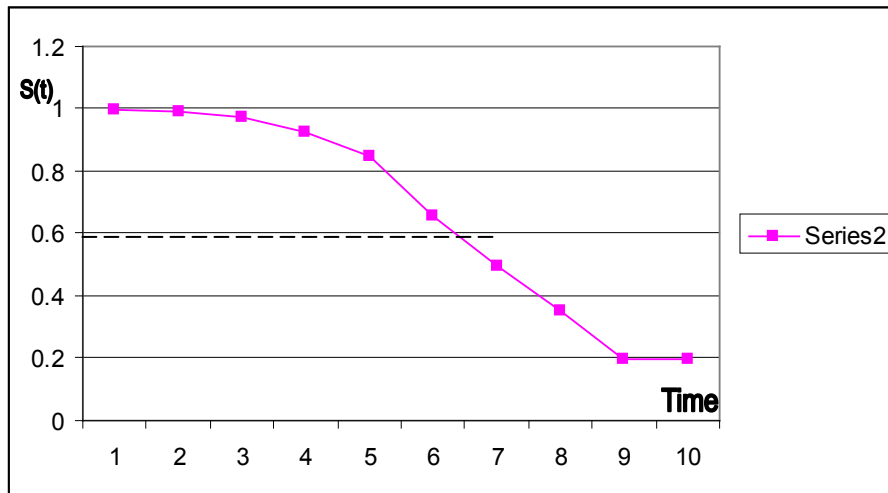**Figure 4.**   Kaplan Meier Curves for Teenager and Adult

Figure 5. Kaplan Meier curves for Antenatal Data

4. Conclusions

From the result we concluded that the probability that pregnant women will continue their antenatal care reduces with time. Moreover, the mean and median survival time being equal to 6 reveals that pregnant women attend their antenatal care six times on average. From the results we also concluded that the antenatal appointment data can best be analyzed parametrically using the Weibull function.

The probability of a pregnant woman attending the antenatal care for time (t) can be modeled as $S(t) = e^{-\lambda t^{\wedge}\alpha}$

The probability that a pregnant woman will stop attending the antenatal care after time t can be modeled as:

$$\Lambda(t) = \lambda\ \alpha\ t^{\alpha-1}$$

Where $\lambda = 0.004$ and $\alpha = 2.89$.

REFERENCES

[1] Bradburn, M. J. et, al (2003). Survival analysis part 1: basic concepts and first analyses. British Journal of cancer 89(2): 232-238.

[2] Collett, D (1994). Modeling survival Data in medical Research. London: Chapman & Hall.

[3] Gross, A. J. and Clark, V. A (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. John Willey & Sons New York.

[4] John Fox (2014). Introduction to Survival Analysis. Lecture notes (online).http://socserv.mcmaster.ca/for/courses/soc761 /survival analysis. pdf.

[5] Ma, Z. and Krings A. W (2008). Survival Analysis Approach To Reliability, Survivability and Prognostics and Health Management. The 2008 IEEE-AIAA Aerospace Conference, Montana, March 1-8, 2008.

[6] Narendranathan et. al. (1993). Modeling the Probability of Leaving Unemployment: competing risks models with flexible baseline hazards. Journal of the Royal Statistical Society, series c, Applied statistics, 42 (1), pp. 63-83.