STATISTICAL ANALYSIS OF FACTORS CONTRIBUTING TO DEATHS OF PATIENTS WITH TUBERCULOSIS

Beki D. O., Elepo T.A., Suleiman Sikiru, Yusuf G.A.

Department of Statistics, Kwara State Polytechnic, Ilorin, Kwara State

Abstract.

Tuberculosis (TB) is a re-emerging infectious disease of international health priority. It is particularly worrisome in Africa, which informed the declaration of public health emergency by the World Health Organisation in 2005. Every year, there are about 8 million new cases and between 2 to 3 million deaths resulting from Tuberculosis despite Tuberculosis is a curable disease. In this study, we carry out a statistical analysis on deaths of the patients with tuberculosis (TB) diseases using UITH, Ilorin as case study. The factors considered for TB status (dead or alive) are sex, age and length of stay. Descriptive statistics and logistics regression was used to analyze the data. The results reveal that patients age-group (eighty-one years plus) are mostly affected by deaths while the patients with less than twenty days on treatment recorded the highest numbers of deaths. The distribution of deaths with respect to sex, age group and length of stay is positively skewed. The odds ratio of 1.16 indicates that the probability of deaths for both sexes is not the same because male patients are more likely to die than female patients. The second model with two-predictor (age and length of stay) provided a statistically significant improvement over the first model with three-predictor (age, sex, and length of stay), the Nagelkerke R^2 indicated that the second model accounted for 87.6% of the total variance in response and the correct prediction rate was about 92.3%. Therefore, we suggest that the second model describes the data better. Lastly, the Hosmer–Lemeshow test with p = .999 indicates that the numbers of deaths are not significantly different from those predicted by the second model and that the overall model fit is good. Finally, the data shows that the age and length of stay on treatment can predict the death of patients with tuberculosis.

Keywords: Tuberculosis; Length of Stay; Status; Age; Sex.

1. Introduction

Tuberculosis (TB) has a long, rich history, dating back as far as Ancient Egypt, with evidence of its presence found in the preserved spines of Egyptian mummies. In the 18th and 19th centuries, a tuberculosis epidemic rampaged throughout Europe and North America, before the German microbiologist Robert Koch discovered the microbial causes of tuberculosis in 1882 *(Comas, et al., 2009)*. Following Koch's discovery, the development of vaccines and effective drug treatment led to the belief that the disease was almost defeated. Indeed, at one point the United Nations predicted that tuberculosis would be eliminated worldwide by 2025. However, in the mid-80s, TB cases began to rise once more in the US and worldwide, so much so that in 1993 the World Health Organization (WHO) declared that TB was a global emergency; the first time that a disease had been labeled as such (NIAID, 2014). **Tuberculosis** is an infectious disease that usually affects the lungs. It is the second greatest killer due to a single infectious agent worldwide, and in 2012, 1.3 million people died from the disease, with 8.6 million falling ill (WHO, 2014). The *Mycobacterium tuberculosis* bacterium causes TB. It is spread through the air when a person with TB (whose lungs are affected) coughs, sneezes, spits, laughs or talks (McIntosh, J., 2015).

According to a study on incidence of Tuberculosis (TB) in Jigawa State, Nigeria, the result shows that there is an increase in the number of patients from 2009 to 2014. Out of the 9590 patients seen in the period, 6538 (68.18 %) are males and3052 (31.82%) are females. A t-test is used to show the significance in the difference between the number of males and females. It is also found that, there is an increase in the number of HIV co-infection over the said years. The study shows the prevalence of TB is highest among the age group of 25 – 34 (30.15%). The commonest type of TB in the population was smear positive pulmonary TB, with 5853 patients (61.03%).

A study on Tuberculosis mortality in Taipei City, Taiwan revealed that the overall mortality rate was 12.3% (249 cases) and the mean age at death was 74 years; 17.3% (43 cases) of all TB deaths were TB-related. Most of the TB-related deaths occurred early (median survival: 20 days), and the patient died of septic shock. Malignancy, liver cirrhosis, renal failure, and miliary and pneumonic radiographic patterns were all independent predictors for all TB deaths. Cavitary, miliary and pneumonic radiographic patterns were all significant predictive factors for TB-related death. Extra pulmonary involvement and liver cirrhosis were also factors contributing to TB-related death (Chou-Han Lin et al., 2014). The majority of TB cases can be cured when the right medication is available and administered correctly. The precise type and length of antibiotic treatment depends on a person's age, overall health, potential resistance to drugs, whether the TB is latent or active, and the location of infection (i.e. the lungs, brain, kidneys). People with latent TB may need just one kind of TB antibiotics, whereas people with active TB (particularly MDR-TB) will often require a prescription of multiple drugs. Antibiotics are usually required to be taken for a relatively long time. The standard length of time for a course of TB antibiotics is about 6 months. All TB medication is toxic to the liver, and although side effects are uncommon, when they do occur, they can be quite serious. Potential side effects should be reported to a health care provider and include: Dark urine, Fever, Jaundice, Loss of appetite, Nausea and vomiting. In 2014, about 80% of reported TB cases occurred in 22 countries. The 6 countries that stand out as having the largest number of incident cases in 2014 were India, Indonesia, Nigeria, Pakistan, People's Republic of China and South Africa. Some countries are experiencing a major decline in cases, while in others the numbers are dropping very slowly. Brazil and China for example, are among the 22 countries with a sustained decline in TB cases over the past 20 years (WHO, 2014).

Tuberculosis is contagious and airborne. It ranks as the second leading cause of death from a single infectious agent, after the human immunodeficiency virus (HIV). 9 million people fell ill with TB in 2013, including 1.1 million cases among people living with HIV. In 2013, 1.5 million people died from TB, including 360 000 among people who were HIV-positive. 510 000 women died from TB in 2013, including 180 000 among women who were HIV-positive. Of the overall TB deaths among HIV-positive people, 50% were among women. TB is one of the top killers of women of reproductive age. An estimated 550 000 children became ill with TB and 80 000 children who were HIV-negative died of TB in 2013. The TB mortality rate has decreased 45% since 1990 (WHO, 2013).

Tuberculosis is a leading infectious disease world-wide and it is among the ten leading causes of death in Nigeria (Stark, 1997). Also TB is a major public health problem in Nigeria, with an estimated 450,000 cases occurring annually of which more than 50% are smear positive (Federal Ministry of Health, 2007). Nigeria has the highest estimated number of new TB cases annually among the African countries (US Embassy in Nigeria, 2013). Also Nigeria ranks the 5th among the 22 high burden countries in the world and total of 86,241 of all forms of TB cases were notified from the 36 states of Nigeria in 2007 (WHO, 2013). In this paper, we draw inferences from the data and empirical observations regarding TB as a major public health challenge using descriptive statistics and logistic regression. The methodology, results and conclusion were discussed in the subsequent sections.

2. Materials and Methods

The study is based on some selected TB patients' data from University of Ilorin Teaching Hospital (UITH), Kwara State, records between 2010 and 2014, with analysis of the observed variables (age, length of stay, sex and their status-dead or alive) in patients with TB who are accessing treatment at the University of Ilorin Teaching Hospital, Kwara State during the period only. Three hundred TB patients were selected using simple random sampling method via their serial numbers in the register with aid of random number table.

The analysis techniques used are descriptive statistics which includes the construction of graphs, charts, and tables, and the calculation of various descriptive measures such as averages, measures of variation, Skewness, percentiles etc. and logistic regression to fit a model for data in which the inference is drawn from. The analysis of the data was performed using Statistical Package for Social Sciences (SPSS) version 21.0. A p-value of less than 0.05 was considered as statistically significant. This section deals with the analysis of data, the results are shown in the table below.

Table 1: Relationship between the number of deaths of patients with TB and their Age Group

Table 1 shows the data grouped into categories according to age group, and the proportion of deaths in each category is given. The proportions of deaths are estimates of the probabilities of death in each category. It suggests that the probability of death increases with the age group; the age group 81+ category 5 has the highest probability of death.

Age group (Years)	of TB Patients	No. of Deaths	oportion of Deaths
<= 20(1)	71	8	0.11
1 - 40(2)	116	25	0.22
1 – 60(3)	77	34	0.44
1 – 80(4)	29	16	0.55
81+(5)	7	4	0.57
Total	300	87	0.29

Table 2: Relationship between the number of deaths and the patients Length of Stay (Los) on TB treatment

Length of Stay (Days)	of TB Patients	o. of Deaths	oportion of Deaths
<= 20(1)	129	46	0.36
21 – 60(2)	89	29	0.33
61 – 100(3)	40	9	0.23
101 – 140(4)	20	2	0.10
141+(5)	22	1	0.05
Total	300	87	0.29

Table 2 shows the data grouped according to length of stay on TB treatment and the proportions of deaths in each category. This suggests that the probability of death decreases with length of stay. Category 1, 0-20 days has the highest probability of death.

Table 3: Coefficient of Skewness from SPSS output

Statistic	Sex	Age	Length of stay
Valid	300	300	300

Missing	0	0	0
Coefficient of Skewness	0.216	0.589	2.068

Table 3 shows that the distribution of data with respect to sex, age and length of stay are positively skewed i.e. stretches more to the right (See Fig.1,2,3 below).

Fig.1 Distribution of TB patients based on Age Group

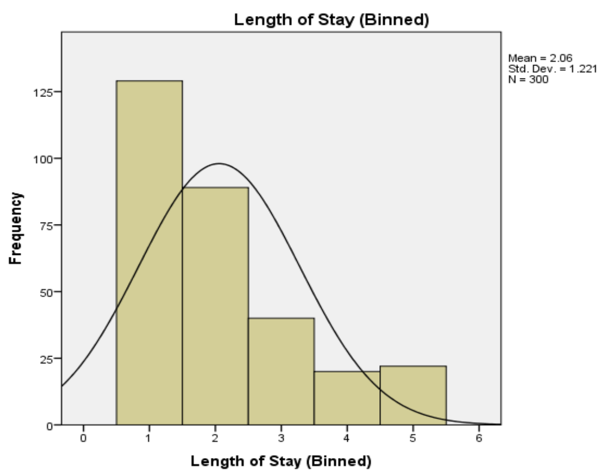

Fig.2 Distribution of patients TB based on their Length of Stay

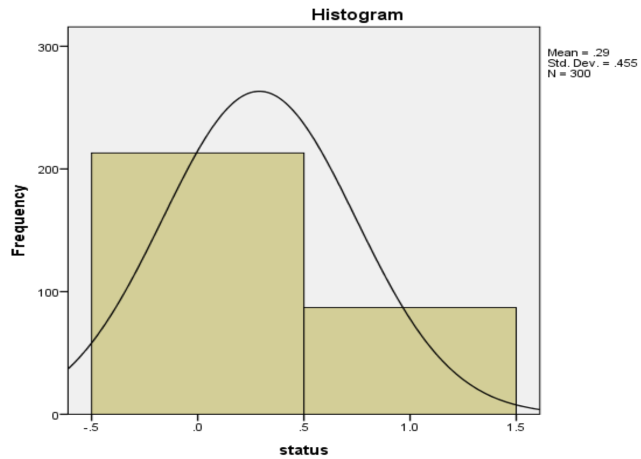

Fig.3 Distribution of patients with TB according to their Sex

Table 4: Classification of Patients Status With Respect To Their Gender

Gender\Status	0 Alive	1 Dead	Total
0 Female	120	46	166
1 Male	93	41	134
Total	213	87	300

From Table 4, the probability of a patient died is 87/300= 0.29, Odds ratio will be used to determine whether this proportion is the same for both sexes. The odds of a male died of TB is 41/93= 0.44, i.e. male is 0.44 times more likely to die than to live; while the odds of a female died of TB are 46/120= 0.38, i.e. female is 0.38 times more likely to die than to live. Hence, the Odds ratio= 0.44/0.38=1.158. Therefore, based on this data, male patients were 1.16 times more likely to die than female patients; hence the probability of death is not the same for both sexes.

3. Model Fit Statistics from SPSS Output

Goodness-of-fit statistics help us to determine whether the model adequately describes the data.

Model Fit - Three-Predictor Model (Sex, Age and Length of Stay)

Table 5: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
	Step	39.041	3	.000
Step 1	Block	39.041	3	.000
	Model	39.041	3	.000

From Table 5, a Chi-Square of 39.041 on 3 df, significant beyond .05. This is a test of the null hypothesis that adding the predictors to the model has not significantly increased our ability to predict the patients' status. Hence, the null hypothesis is rejected ($p < .05$). Hence, the model is significant.

Table 6: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	322.250[a]	.122	.174

From Table 6, the -2 Log Likelihood statistic is 322.250. This statistic measures how poorly the model predicts the patients' status -- the smaller the statistic the better the model. The Cox and Snell R^2 can be interpreted like R^2 in a multiple regression, but cannot reach a maximum value of 1. The Nagelkerke R^2 can reach a maximum of 1. Nagelkerke R^2 is preferred; this suggests that the model accounts for 17.4% of variance of response.

Table 7: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.998	8	.265

From Table 7, Hosmer and Lemeshow Test statistic has a $p = 0.265$ which means that it is not statistically significant at $\alpha=0.05$ and therefore our model is a good fit.

Table 8: Estimates of the three variables model

	B	S.E.	Wald	df	Sig.	Exp(B)

	sex(1)	.148	.273	.292	1	589*	1.159
1[a]	age	.035	.007	23.275	1	.000	1.036
	lengthofstay	-.009	.003	7.041	1	.008	.991
	Constant	-1.986	.381	27.166	1	.000	.137

ri variable (s) entered in column 1 are sex, age, length of stay.

From Table 8, the three-predictor model shows that the logistics regression equation is

$$In(odds) = -1.986 + 0.035Age - 0.009Length\ of\ \ stay + 0.148Sex(1)$$

The variables in the equation include Exp (B). This is better known as the odds ratio predicted by the model.

i. For gender, males are 1.16 times more likely to die than females patients.

ii. For age, patients are 1.04 as likely to die at their age as to be alive.

iii. For length of stay, the probability of patient dying is contingent on length of stay.

Long length of stay is less likely to die. However, sex (1) is not statistically significant to the model at α=0.05, then it will be removed from the model and the new model will be fitted.

Table 9:Classification Table

Observed			Predicted		
			status		Percentage Correct
			0 Alive	1 Dead	
Step 1	Status	0 Alive	194	19	91.1
		1 Dead	67	20	23.0
	Overall Percentage				71.3

The cut value is .500

The classification Table 9 shows that three-predictor model correctly classify 20 / 87 = 23% of status as deaths. This is known as the sensitivity of prediction, i.e. the percentage of occurrences correctly predicted. Also, it correctly classifies 194 / 213 = 91.1% of status as alive. This is known as the specificity of prediction, i.e. the percentage of non-occurrences correctly predicted. Overall correct predictions were 214 out of 300 times, for an overall success rate of 71.3%.

Model Fit - Two-Predictor Model (Age and Length of Stay)

Table 10:Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	76.359[a]	.613	.876

From Table 10, the R^2 statistics have increased and -2 Log Likelihood statistic has dropped from 322.250 to 76.359 which indicated an improvement over the three-predictor model. Nagelkerke R^2 suggests that the new model accounts for 87.6% of variance of response

Table 11:Hosmer and Lemeshow test

	Chi-square	df	Sig.
	.738	8	.999

From Table 11, Hosmer and Lemeshow Test statistic has a $p = 0.999$ which is not statistically significant at α=0.05; the model is quite a good fit and better than three-predictor model. This indicates that the numbers of deaths are not significantly different from those predicted by the model.

Table 12: Estimates of the Two variable model

		B	S.E.	Wald	df	Sig.	Exp(B)
	age	.035	.007	23.270	1	.000	1.036
Step 1[a]	lengthofstay	-.009	.003	7.049	1	.008	.991
	Constant	-1.918	.358	28.751	1	.000	.147

:iVariable(s) entered in column 1: age, length of stay.

For the two-predictor model, the variables in the equation output shows us that the logistic regression equation is

$$In(odds) = -1.918 + 0.035 Age - 0.009 Length\ of\ stay$$

Table 13:Classification Table[a]

		Observed	Predicted		
			Status		Percentage Correct
			0 Alive	1 Dead	
Step 1	status	0 Alive	202	11	94.8
		1Dead	12	75	86.2
	Overall Percentage				92.3

e cut value is .500

Table 13 shows that two-predictor model correctly classify 75 / 87 = 86.2% of status as deaths. Also, it correctly classifies 202 / 213 = 94.8% of status as alive. Overall correct predictions were 277 out of 300 times, for an overall success rate of 92.3%. For three-predictor model, a false positive is 19 / 39 = 49%.i.e. predicting that the event would occur when it did not and false negative is 67 / 261 = 26%.i.e. predicting that the event would not occur when it did occur. Hence, this model cannot adequately describe the data due to higher error rate. For two-predictor model, a false positive is 11 / 86 = 13%.i.e. predicting that the event would occur when it did not. A false negative is 12 / 214 = 6%.i.e. predicting that the event would not occur when it did occur. Therefore, this model can adequately describe the data because the error rate is lower.

4. Summary of Findings

From the above analysis, the followings are summary of findings. The Table 1 suggests that the probability of deaths increases with the age-group and the 81 plus is the age-group which is mostly affected while Table 2 shows that the probability of deaths decreases with length of stay in which less 20 days recorded the highest proportion of deaths. This is in line with the claim that about 60% of death during TB treatment was in the first two months (Daniel et al, 2007).

The distribution of data with respect to sex, age and length of stay skewed to the right. This means the data is not symmetric. The Odds ratio indicates that the male patients were 1.16 times more likely to die than female patients; hence the probability of death is not the same for both sexes. The three-predictor model accounts for 17.4% of the total variance and correct prediction rate was about 71.3% with higher error rate in classification. Also, the result shows that sex (1) is not statistically significant to the model. Hence, it cannot be used to predict deaths. After removing the sex (1) which is not significant to the first model, the two-predictor model (second model) accounts for 87.6% of the total variance and correct prediction rate was about 92.3%. The error rate in classification is very low, therefore, we suggests that the second model can describe the data adequately.

5. Conclusion

In conclusion, the patients eighty years and above are mostly affected by death while patients with less than twenty days on treatment recorded the highest numbers of death. The distribution of deaths with respect to sex, age group and length of stay is positively skewed. The odds ratio of 1.16 indicates that the probability of dying for both sexes is not the same because male patients are more likely to die than female patients. From logistic regression analysis, the two-predictor model (age and length of stay) provided a statistically significant improvement over the first model (three-predictor model). The Nagelkerke R^2 indicated that the two-predictor model accounts for 87.6% of the total variance and the correct prediction rate was about 92.3%. Therefore, we suggest that the two-predictor model describes the data. Lastly, the Hosmer–Lemeshow test indicates that the numbers of deaths are not significantly different from those predicted by the model and that the overall model fit is good. Therefore, we concluded that the age and length of stay can adequately predict the death of patients with TB.

References

Baba and Hincal. (2015). Statistical Analysis of Tuberculosis in Jigawa State, North-Western Nigeria

Barry R. (1994). Tuberculosis: Pathogenesis, protection, and control. Washington, D.C.: ASM Press, USA

Caroline Hannaway (2008). Biomedicine in the twentieth century: practices, policies, and politics. Amsterdam, Holland

Chou-Han Lin et al.,(2014). Tuberculosis mortality: patient characteristics and causes. Department of Internal Medicine, Far Eastern Memorial Hospital, Banciao Dist., New Taipei City, Taiwan . BMC Infectious Diseases 14:5

Comas. I, Gagneux, S(2009). The past and future of tuberculosis research. *Manchester, Marianne, ed. UK.*

Corbett E. L., Watt C. J., et al.,(2003). The Growing Burden of Tuberculosis: Global Trends and Interactions with the HIV Epidemic;163:1009–21.

Daniel OJ, Salako A.A, et al., (2007). Trend in Tuberculosis mortality in Nigeria. Dept. of community medicine and primary care. Olabisi Onabanjo university teaching hospital, sagamu, Ogun state, Nigeria

Dye C, Scheele S, et al., (1999). Consensus statement global burden of tuberculosis; Estimated incidence, prevalence and immortality by country. WHOs Global Surveillance and Monitoring Project. JAMA 282: 677-686.

Federal Ministry of Health (2007). National Guidelines for TB and HIV/AIDS Treatment and Care in Adolescents and Adults. Abuja, Nigeria. Pp. 1-4.

Harries A. D., Hargreaves N. J., et al.,(2001). Tuberculosis in Malawi: Nationwide Case-Finding and Treatment Outcomes. International Journal of Tuberculosis and Lung Disease.6:424–31.

Konomi N, Lebwohl E, et al., (2002). Detection of Mycobacterial DNA in Andean Mummies. *J Clin Microbiol 40 (12): 4738–40.*

Lawn, S.D, Zumla, AI (2011). Tuberculosis. Lancet 378 (9785): 57–72.

Lucas S. B., Hounnou A., et al., (1993). The Mortality and Pathology of HIV Infection in a West African City. AIDS. 7:1569–79

McIntosh, J. (2015). Tuberculosis: Causes, Symptoms and Treatments. Medical News Today

Omoleke S.A. (2012). An analysis of tuberculosis in developing and developed world: Nigeria and UK as a case study, Journal of Public Health and Epidemiology Vol. 4(6), pp. 150-155

Pearce-Duvet J (2006). The origin of human pathogens: evaluating the role of agriculture and domestic animals in the evolution of human disease. Biol Rev Camb Philos Soc 81 (3): 369–82.

Persson, Sheryl (2010). Smallpox, Syphilis and Salvation: Medical Breakthroughs That Changed the World. p. 141

Rothschild B.M, Martin L.D, Lev G, et al. (2001). Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin. Infect. Dis. 33 (3): 305–11.*

Sledzik, P.S. and Nicholas Bellantoni (1994). Bioarcheological and biocultural evidence for the New England vampire folk belief, *American Journal of Physical Anthropology 94 (2): 269–274.*

Stark J.R. (1997). Tuberculosis: An old disease but a new threat to the mother, fetus and neonate. Clinics in perinatology. 24: 107-127

US Embassy in Nigeria (2013). Nigeria Tuberculosis Fact Sheet. United States Embassy in Nigeria, Diplomatic Drive Central Area Abuja, FCT, Nigeria

World Health Organisation (2013). Global Tuberculosis Control: A Short Update to 2009 Report. Geneva, Switzerland.

World Health Organization (2014). Report of global tuberculosis control: Surveillance, planning and financing. World Health Organization, Geneva, Switzerland