Journal of Science (2013) Vol. 1 No. 1: pp. 1-10
Received: 13 August 2013 / Accepted: 15 November 2013 / Published online: 20 November 2013
© Academy of Science and Social Science 2013

Tuberculosis: A Study of Patients in Nigeria Using Binary Logit Models

Adebowale O. Adejumo[1], Elizabeth D. Bright[1], Ifeyinwa V. Omekam[1], Remi E. Popoola[1], Adesupo A. Akinrefon[2], Oluwole A. Odetunmibi[3], and Adetunji A. Ademola[4]

[1]*Department of Statistics, University of Ilorin, Nigeria*

[2]*Department of Statistics and Operations Research, Modibbo Adama University of Technology, Nigeria*

[3]*Department of Computer and Information Sciences, Covenant University, Nigeria*

[4]*Department of Mathematics and Statistics, Federal Polytechnic, Nigeria*

*Corresponding Author: Adebowale O. Adejumo

*Corresponding Author's Email: aodejumo@unilorin.edu.ng

Abstract

Purpose: This study aims to examine the influence of age, state, and year on gender of Tuberculosis patients using binary logit modeling.

Design/Methodology/Approach: Binary logit models were computed using data sets of registered Tuberculosis patients from 2006 to 2009.

Research Findings: Results suggest that males face higher risks for Tuberculosis as compared to females in all age groups, states, and years. In addition, risk variation was observed in age groups, states, and years.

Research Limitations/Implications: This study is limited in its analysis to only data sets of registered Tuberculosis patients from 2006 to 2009.

Originality/Value: This study contributes to a better understanding of Tuberculosis patients in Nigeria in terms of age, state, year, and gender.

Keywords: Tuberculosis, patients, binary logit models, Nigeria.

Manuscript Category: Research paper

Introduction

In 1993, the World Health Organization (WHO) took an unprecedented step and declared Tuberculosis (TB) a global emergency, the first of such designated ever made by the organization. TB kills approximately two million people each year. One individual becomes infected with the bacteria that cause TB every second. One third of the world's population is infected with the bacteria. As many as one in ten of those infected will develop active symptoms of TB at some point in their lives.

Most studies on TB are carried out in the Western context. This study seeks to contribute to the understanding of TB from the African context. In particular, this study aims to understand the influence of age, state, and year on gender of TB patients in Nigeria using binary logit modeling. Gender is used as the response variable while age, state, and year are used as explanatory (or independent) variables. The data used in this study was collected by the Ministry of Health in Leprosy and the Tuberculosis Clinic in Kwara. The data consists of 16,514 cases of registered TB patients, out of which 9,638 were males and 6,876 were females. The data was classified in terms of age, states, and years. Trends of new sputum pulmonary TB cases were registered over a period of four years (2006-2009).

The remainder of this paper is organized as follows. The existing literature on TB is first reviewed. This is followed by a discussion on the research methodology and data analysis and results. Finally, some concluding remarks are made based on the results from the data analysis.

Literature Review

TB is a disease that has existed for thousands of years. It was derived from the world "tubercular", which refers to small scars in tissues (i.e. small lumps), and was first used in 1839. In the 19th century, an epidemic of the disease occurred in Europe and the United States. This led to a plethora of studies on the causes of and cures for TB (WHO, 2008).

Causes of Tuberculosis

Multiple factors contribute to the global increase in TB infection. The human immunodeficiency virus (HIV) that causes acquired immune deficiency syndrome (AIDS) poses the greatest risk to activate and progress TB. People with HIV have a weakened immune system that increases their susceptibility to TB, and in these people, TB often progresses rapidly from the primary to secondary stage. This is observed in statistics showing that the increase in TB incidence is highest in Africa and Asia, which consists of the highest number of people infected with HIV (WHO, 2009a; 2009b).

TB resurgence is also contributed by the failure of patients to complete the full six months of antibiotics therapy required to cure the disease. Many people stop taking antibiotic when they begin to feel healthier. However, successful treatment of TB requires therapy beyond the period of obvious symptoms. When patients fail to follow prescribed treatments, they may become actively infectious, spreading the disease to others. Furthermore, failure to

complete full therapy can cause emergence of TB bacterial strains with acquired drug resistance, thereby further complicating treatment by increasing the length and cost of therapy.

Other factors such as migration, international air travel, and tourism also contribute to the widespread of TB. The extreme difficulty of screening immigrants and travelers for TB allows the disease to cross international borders easily. The substantial increase in homelessness and the related circumstances of poverty, overcrowding, and malnutrition also contributed to the increased incidence of TB in the United States and other industrialized countries during the early 1990s. Industrialized nations with good public health systems have been able to control the recent TB resurgence. However, curbing the spread of TB on a global scale will require ongoing international efforts. In the future, combating TB throughout the world will require advances in molecular biology, in-depth research on the genetics of TB in order to understand drug resistance, and the continuous development of new drugs, as well as the prospect of synthesizing additional vaccines (Atun et al., 2005; Rios et al., 2000, Bacaër et al., 2008).

Cures for Tuberculosis

Several cures for TB have been outlined in the literature. American physician Edward Trudeau was affected by the disease twice, in 1873 and 1876. When he thought he was dying, he traveled to Saranac Lake in the Adirondack Mountains of New York to spend his final days. When he found his symptoms eventually cure, he attributed his recovery to the fresh air of the mountains. In 1885, Trudeau built the first American sanatorium, which later became a model for many sanatoriums. That became the main stay of TB treatment in the late 19th century and 20th century. By 1930, the United States had 600 sanatoriums with a total of 84,000 beds. Trudeau also established the Trudeau laboratory, which was responsible for training most physicians versed in the treatment of TB.

Early in the 19th century, TB was considered a refined disease – that is, one that affected artistic, morally superior individuals. However, as the epidemic continued and claimed a larger circle of people, often the poor and the disadvantaged. The victims themselves were often blamed, and in the absence of scientific knowledge, TB was attributed to a person's lifestyle. Scientific pursuit of the true nature of TB continued, and in 1882, German physician Robert Koch discovered the bacteria that caused TB and demonstrated the presence of the bacteria and how it was transmitted using simple but precise observations and experiments (WHO, 2008, 2009a, 2009b).

In Paris, French bacteriologists Albert Calmette and Camille Guerin worked with a virulent strain of bovine (low) tubercle bacillus at the Pasteur institute. In 1924, they prepared the BCG vaccine in hopes of protecting the world against TB. The vaccine was administered to a new born child who was at the high risk of developing TB. The vaccine was successful and the child never contacted the disease.

In 1944, American micro biologist Selman Waksman isolated Streptomycin from a fungus, Streptomycin Lavendula, heralding the beginning of modern antibiotic therapy for TB. Sources of drug therapy and declining rates of disease incidence and mortality over the next 30 years instilled a sense of confidence in public health officials that TB could be conquered. As antibiotic therapy became the primary treatment, mortality rates from TB decreased significantly. Deaths from TB in the United States dropped from 188 per 100,000 in 1904 to about 1 per 100,000 in

1980. From 1953 to 1984, the average annual decline in cases was about 3 percent per year. As a result, funding for public health programs in the United States, including those for the prevention and treatment of TB, was drastically curtailed in 1980s. (Atun et al., 2005; Rios et al., 2000, Bacaër et al., 2008).

Research Methodology

Most studies distinguish between explanatory and response variables, whereby modeling effects of explanatory variables on response variables are more important than modeling relationships among explanatory variables. The goal of logit modeling is to find the best fitting and most parsimonious model to describe the relationship between an outcome variable and a set of independent variables. The main difference between logit and linear regression models is that the outcome variable in logit is binary (Peng et al., 2002). Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictors (or explanatory variables). More often, the response is not a numerical value but a designation of one of two possible outcomes: alive or dead, success or failure, yes or no (Bayaga, 2010).

Log linear models contain the same structure as logit models for association between explanatory and response variables and they contain the most generalized interaction term for relationship among explanatory variables. Logit models are used to model relationships between a dependent variable (Y) and one or more independent variable (X). The dependent variable is a discrete variable that represents a choice, or category, from asset of mutually exclusive choices or categories.

The observations on Y are assumed to have been randomly sampled from the population of interest (even for stratified sample or choice-based samples). Y is caused by or associated with the X's and X's are determined by influences (variables) outside of the model. There is uncertainty in the relation between Y and the X's as reflected by a scattering of observations around the functional relationship. The distribution of error term must be assessed to determine if a selected model is appropriate (see Agresti, 2002; Bayaga, 2010; Peng et al., 2002; Hosmer and Lemeshow, 2000).

Considering a four way classification in which gender (G) is a binary response variable on explanatory factors: age (A), state (S), and year (Y).

The logit model –

$$Logit\left[P(G=1)\right]=\log\left(\frac{P(G=1)}{1-P(G=1)}\right)=\log\left(\frac{P(G=1)}{P(G=2)}\right)=\log\left(\frac{\pi_{hij1}}{\pi_{hij2}}\right)=\alpha+\beta_h^Y+\beta_i^S+\beta_j^A \qquad (1)$$

– contain main effect term (β_h^Y ; β_i^S *and* β_j^A) but no interaction term, π_{hij1} represents the proportion of male category while π_{hij2} represents the female category. The additivity on the logit scale is the standard definition of "no interaction" for categorical variables. This model corresponds to the log linear model (2) that contains the full interaction term among the factors and association between each factor.

$$\log(m_{ijkl}) = \mu + \lambda_h^Y + \lambda_i^S + \lambda_j^A + \lambda_g^G + \lambda_{hi}^{YS} + \lambda_{hj}^{YA} + \lambda_{hg}^{YG} + \lambda_{ij}^{SA} + \lambda_{ig}^{SG} + \lambda_{jg}^{AG} + \lambda_{hij}^{YSA} \quad (2)$$

One can check that by equating

$$\beta_h^Y = \lambda_{h1}^{YG} - \lambda_{h2}^{YG},\ \beta_i^S = \lambda_{i1}^{SG} - \lambda_{i2}^{SG}, \text{ and } \beta_j^A = \lambda_{j1}^{AG} - \lambda_{j2}^{AG} \quad (3)$$

Estimating the parameters in the model can be done using any of the following methods for the Maximum Likelihood (ML) estimation; Newton-Raphson, Fishers Scoring or Iterative Proportional Fitting (IPF), since the logit model also falls into the class of Generalized Linear Model (GLM). In this study, Fishers scoring iterative method will be used (Adejumo, 2005; Adejumo and Arabi, 2013; Christensen, 1997). The equation of the Fisher-scoring algorithm is:

$$(X^1 W^{(k)} X)\beta^{(k+1)} = X^1 W^{(k)} Z^{(k)} \quad (4)$$

This is the likelihood equation of a generalized linear model with the response vector $Z^{(k)}$ and a random error covariate matrix $(W^{(k)})^{-1}$. If rank(X) = p holds, we obtain the ML estimate β as the limit of $\beta^{(k+1)} = (X^1 W^{(k)} X)^{-1} X^1 W^{(k)} Z^{(k)}$.

An algorithm was written in R programme to evaluate the process described and the parameters, log-likelihood ratio statistic (G^2) as well as the Akaike Information Criteria (AIC) are obtained for logit model stated in equation (1).

Data Analysis and Results

The period of the data used was from 2006 to 2009. It consists of a total of 16,514 cases of registered TB patients, out of which 9,638 are male and 6,876 are female. This was further classified in terms of their age groups, states and years.

Table 1 (see Appendix) presents the parameter estimates for the logit model in (1) with three factors. Each estimated coefficient is the expected change in the *log* odds of age, state, and year. The corresponding odds ratio estimation from the fitted logit model are shown in Table 2 (see Appendix). A summary of results from the data analysis is as follows:

The Effect of Year on Gender Using Year 2006 as Reference Point

β_{2007}-β_{2006} = 0.31838-0.95806= -0.63968

β_{2008}-β_{2006} = 0.43524-0.95806= -0.52285

β_{2009}-β_{2006}= 0.24044-0.95806= -0.75362

The Effect of Age Group on Gender Using Age Group 1

β_{Ag2}-β_{Ag1} = .0.13966+0.72311=0.58345

β_{Ag3}-β_{Ag11} = .0.17643+0.72311=0.54668

β_{Ag4}-β_{Ag1} = .0.03803+0.72311=0.68508

β_{Ag5}-β_{Ag1} = .0.0995+0.72311=0.62361

β_{Ag6}-β_{Ag1} = .0.10107+0.72311=0.62204

β_{Ag6}-β_{Ag1} = .0.16841+0.72311=0.5547

The Effect of State on Gender Using Benue State

β_{fct}-β_{Benue} = -0.29477

β_{Kogi}-β_{Benue} =0.20297

β_{Kwara}-β_{Benue} =-0.02428

$\beta_{Nasarawa}$-β_{Benue} =-0.19038

β_{Niger}-β_{Benue} =-0.49305

$\beta_{Plateau}$-β_{Benue} =-0.29039

Conclusion

Four variables were investigated in this study: age, state, year, and gender. The odds ratio when there is no interaction was examined. The estimated odds that males have TB is higher than females in all age groups, states, and from 2006 to 2009. Also from further test, Plateau state was found to have the highest estimated odds within all age groups. Age group 1 (15-24 years) has the highest estimated odds. In north central zones, 2006 has the highest estimated odds (exp(0.95806)=2.607).

For two way interaction, the estimated odds that males have TB is higher within all age groups. Males have higher estimated odds in all north central zones except for Niger. Also, males have higher estimated odds within all age groups. The highest estimated odds for TB is in 2007 in all age groups, whereas 2008 has the highest estimated odds in all north central zones except Benue and Kogi. When three way interaction is considered, males have higher estimated odds in all age groups and in all north central zones except for Nassarawa. Considering four way interaction, females have higher estimated odds of TB in age groups 3,5,6 and 7 than males. In the north central zones, males have higher estimated odds, except for Niger. Also, males have higher estimated odds in 2006, 2007, 2008 and 2009.

This study also shows that the effect of having TB for a given year has a risk of 52% higher in 2004 than 2003, 59% higher in 2008, and 49% higher in 2009 than 2006 (see Table 4). Furthermore, considering the effect of age group (see Table 5), the risk is 1.7922 higher in age group 2 than age group 1, 1.72751 times higher in age group 3 than age group 1, 1.98393 times higher in age group 4 than age group1, 1.86505 times higher in age group 5 than age group 1, 1.86272 times higher in age group 6 than age group 1 and 1.74142 times higher in age group 7 than age group 1. Lastly, considering the effect of states (see Table 6), the risk is 74% higher in Fct than Benue, 122% higher in Kogi than Benue, 97% higher in Kwara than Benue, 83% higher in Nassarawa than Benue, 61% higher in Niger than Benue and 75% higher in Plateau than in Benue state. Hence, this study concludes that the risk of having TB, varies from year to year, age group to age group and state to state.

References

Adejumo, A. O. (2005). *Modelling generalized linear (loglinear) model for raters agreement measures*. Frankfurt am Main: Peter Lang.

Adejumo, A. O. and Arabi, F. J. (2013). A study of HIV sero-prevalence surveillance survey data using loglinear model. *Global Journal of Pure and Applied Mathematics*. Vol. 9 No. 1, pp. 73-81.

Agresti, A. (2002). *Categorical data analysis*. New York: John Wiley and Sons.

Atun, R. A, Samyshkin, Y. A, and Drobniewski, F. (2005). Seasonal variation and hospital utilization for tuberculosis in Russia: hospitals as social care institutions. *European Journal of Public Health*, Vol. 15, pp. 350-354.

Bacaër, N., Ouifki, R., and Pretorius, C. (2008). Modeling the joint epidemics of TB and HIV in a South African township. *Journal of Mathematical Biology*, Vol. 57, pp. 557-593.

Bayaga, A. (2010). Multinomial logistic regression: usage and application in risk analysis. *Journal of Applied Quantitative Methods*, Vol. 5, No. 2, pp. 288-297.

Christensen, R. (1997). *Loglinear models and logistics regression*. New York: Springer-Verlag.

Hosmer, D. W., and Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.

Peng, C. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Education Research*, Vol. 96 No. 1, pp.3-14.

Rios, M., Garcia, J. M, Sanchez, J. A, et al. (2000). A statistical analysis of the seasonality in pulmonary tuberculosis. *European Journal of Epidemiology*, Vol. 16 No. 5, pp. 483-488.

World Health Organization. (2008). Global tuberculosis control: surveillance, planning, financing. *WHO Report 2008*, Geneva.

World Health Organization. (2009a). Tuberculosis control in the South East Asia region. *WHO Report 2009*, Geneva.

World Health Organization. (2009b). Global tuberculosis control 2009: epidemiology, strategy, financing. *WHO Report 2009*, Geneva.

Appendix

Table 1 Parameter Estimates for the Fitted Logit Model

Coefficient	Estimates	Standard Error	Z-value	P-value
intercept	0.23421	0.07271	3.221	0.00128 **
β^{Y}_{2007}	0.31838	0.05154	6.177	6.53e-10 ***
β^{Y}_{2008}	0.43524	0.04854	8.966	< 2e-16 ***
β^{Y}_{2009}	0.20444	0.04893	4.178	2.94e-05 ***
β^{A}_{ag2}	-0.13966	0.07629	-1.831	0.06716
β^{A}_{ag3}	-0.17643	0.07117	-2.479	0.01317 *
β^{A}_{ag4}	-0.03803	0.07170	-0.530	0.59585
β^{A}_{ag5}	-0.09951	0.06673	-1.491	0.13592
β^{A}_{ag6}	-0.10107	0.06864	-1.472	0.14093
β^{A}_{ag7}	-0.16841	0.07054	-2.388	0.01696 *
β^{S}_{fct}	-0.07679	0.05332	-1.440	0.14984
β^{S}_{kogi}	0.42095	0.06345	6.634	3.27e-11 ***
β^{S}_{kwara}	0.19370	0.06522	2.970	0.00298 **
$\beta^{S}_{nassarawa}$	0.02760	0.06409	0.431	0.66667
β^{S}_{niger}	-0.27507	0.06315	-4.356	1.33e-05 ***
β^{S}_{plateu}	-0.07241	0.06952	-1.042	0.29759

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Likelihood Ratio (G^2) = 21356 on 371 degrees of freedom

AIC = 21388

Table 2 Estimates of Effect Factors' Odds Ratios

Effect Factors	Estimates	Odds Ratio Estimates
Year 2007	0.31838	Exp(0.31838) =1.3749
Year 2008	0.43524	Exp(0.43524) =1.54533
Year 2009	0.20444	Exp(0.20444) =1.22681
Age group 2	-0.13966	Exp(-0.13966)=0.86965
Age group 3	-0.17643	Exp(-0.17643)=0.83826
Age group 4	-0.03803	Exp(-0.03803)=0.96268
Age group 5	-0.09951	Exp(-0.09951)=0.90528
Age group 6	-0.10107	Exp(-101017) =0.90387
Age group 7	-0.16841	Exp(-0.16841)=0.845007
Fct	-0.07679	Exp(-0.07679)=0.92608
Kogi	0.42095	Exp(0.42095) =1.52341
Kwara	0.19370	Exp(0.19370) =1.21373
Nassarawa	0.02760	Exp(0.02760) =1.02798
Niger	-0.27507	Exp(-0.27507 =0.75952
Plateau	-0.07241	Exp(-0.07241)=0.93015

Table 3 Estimates of Reference Point for Each Factor

Factors	Estimates
Year 2006	0.31838+0.43524+0.2044=0.95806
Age group 1	-0.13966-0.17643-0.03803-0.09951-0.10107-0.16841=0.72311
Benue State	-0.07679+0.42095+0.19370+0.02760-.27507-0.07241=0.21798

Table 4 (β_{yr}-β_{06}) and Estimated exp (β_{yr}-β_{06})

	2007	2008	2009
Estimate	0.31838	0.43524	0.24044
(β_{yr}-β_{06})	-0.63968	-0.52282	-0.75362
Exp(β_{yr}-β_{06})	0.52746	0.59285	0.47066

Table 5 (β_{Ag}-β_{Ag1}) and Estimated exp(β_{Ag}-β_{Ag1})

	Age Group 2	**Age Group 3**	**Age Group 4**	**Age Group 5**	**Age Group 6**	**Age Group 7**
Estimate	-0.13966	-0.17643	-0.03803	-0.09951	-0.10107	-0.16841
(β_{Ag}-β_{Ag1})	0.58345	0.54668	0.68508	0.62361	0.62204	0.5547
Exp(β_{Ag}-β_{Ag1})	1.7922	1.72751	1.98393	1.86505	1.86272	1.74142

Table 6 (β_{State}-β_{Benue}) and Estimated exp(β_{State}-β_{Benue})

	Fct	**Kogi**	**Kwara**	**Nasarawa**	**Niger**	**Plateau**
Estimate	-0.07679	0.42095	0.19370	0.02760	-0.27507	-0.07241
(β_{State}-β_{Benue})	-0.29477	0.20297	-0.02428	-0.19038	-0.49305	-0.29039
exp((β_{State}-βBenue)	0.74470	1.22504	0.97601	0.82664	0.61076	0.74797