INVESTIGATING THE DYNAMICS OF TUBERCULOSIS AND HUMAN IMMUNODEFICIENCY VIRUS AMONG TUBERCULOSIS PATIENTS IN FEDERAL MEDICAL CENTRE KEFFI USING LOGISTIC REGRESSION MODEL

Adenomon M. O.

. Department of Statistics, Nasarawa State University, Keffi, Nigeria & NSUK-LISA Stat Lab, Nasarawa State University, Keffi, Nigeria
International Association of Statistical Computing (IASC) African Members Group
Foundation of Laboratory for Econometrics and Applied Statistics of Nigeria

Achi M. E.

Department of Statistics, Nasarawa State University, Keffi, Nigeria & NSUK-LISA Stat Lab, Nasarawa State University, Keffi, Nigeria

Abstract
Tuberculosis (TB), remains the leading cause of death among people with HIV and one of the top 10 leading causes of death from a single infection worldwide (WHO, 2020). In 2018, Nigeria was listed as the first in Africa and sixth among the 30 countries of the world with highest burden of TB. Over time study has shown that the emergency and spread of drug resistance TB and high burden of HIV/AIDs has contributed to the complication and wide spread of TB in Nigeria. This study investigated the dynamics of TB and HIV among TB patients in Federal Medical Center Keffi, Nigeria using Logistic regression analysis. To achieve the aim of this study, dataset on recorded patients with Tuberculosis cases were collected from the DOT center, Federal Medical Center keffi, Nigeria. The data consisted of 1606 registered TB cases recorded between January 2015 to December 2020. Out of the 1606 TB cases examined, 693 representing 43.2% were females, while 913 representing 56.8% were male gender. Findings from the study show that out of the 1606 TB cases recorded, 308(19.2%) were diagnosed with extra pulmonary TB, while the remaining 1298(80.8%) cases were diagnosed to have pulmonary TB. Also, out of the 1606 TB cases recorded in the study, 1166(72.6%) tested HIV negative, while the remaining 440(27.4%) tested HIV positive. The findings obtained from the logistic regression analysis, for HIV model it revealed that age is statistically insignificant (p-value=0.431>0.05). But according to the result from the odds ratio test for HIV, older TB patients in age are 0.2% more likely to be infected by HIV compared to younger TB patients in age. Also, from the odds ratio test carried out, we found out that male TB patients has 86% more chances of been affected by HIV than the female TB patients. The result also indicated that patients with pulmonary TB cases have a 62% greater odds of developing HIV infection compared to does with extra-pulmonary TB. Classification test of logistic regression for both HIV and TB revealed to us about 72.60% and 80.82% level of correct classification, which implies a high level of agreement for our model. Finally, the result revealed that HIV and TB affects each other in a bi-directional way.

Keywords: HIV, TB, Logistic Regression, Goodness-of-fit, ROC.

1 Introduction

Tuberculosis (TB) remains a global public health problem and one of the top ten leading causes of death, worldwide, with developing countries bearing the highest burden (WHO, 2020). In 2018, Nigeria was listed as first in Africa and sixth among the 30 countries of the world with the highest TB burden (WHO, 2020). Unfortunately, the problem of TB in Nigeria has been complicated by the emergence and spread of drug resistant TB and a high burden of HIV/AIDS (NTBLCP, 2017; WHO, 2018). The problem of TB is worsened when there is also a high burden of HIV infections, as people with HIV are more likely to develop active TB. According to WHO reports, an estimated 63,000 Nigerians living with HIV/AIDS develop TB, while about 39,000 die from the disease, each year (WHO, 2018).

The main goal of this research is to determine the dynamics in TB/HIV cases among TB patients registered at the Direct Observed Treatment Short Course (DOTS), Federal Medical Center Keffi, Nigeria and to estimate the probability. This study has the potential to create improvement in TB/HIV services in Nigeria. Logistic regression analysis would be used to assess the relationship between dependent and independent variables. One of the reasons why we choose logistic regression in this study, is that the dependent (response) variable is categorical in nature.

According to Njepuome and Odume (2009) who examined the impact of HIV syndromes in the treatment of TB cases in Gombe state, Nigeria, using 300 patients with HIV and TB. The study revealed no significant difference in the mean age of male and female. The study also revealed that the death rate among dually infected patients was higher compared to patient with only HIV negative status. Ngowi, et al. (2008), in their study, reported that patients with low immunity due to HIV are more likely to acquire tuberculosis in an area with high tuberculosis prevalence.

2 Model Specification

2.1 Logistic regression

The concept of logistic regression requires the dependent variable should be binary (either 1 and 0 or Yes or No) that is nominal (Zar, 2014) such model has been applied in Awodiji et al., (2023). The logistic regression does not require the assumption of linearity, normality of error terms and homescedastic be fulfilled unlike in Ordinary Least squares (OLS). In this study, HIV status is the dependent variable (Yes=1 and No=0). The model is given by

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \quad (1)$$

Where β_0 is the intercept while β_1, β_2 are the slopes.

The odds ratio is given as $odds\,ratio = \ell^{\beta_1}$

The rational for the use of logistic regression in this study is because the dependent variable is nominal while the independent variables are mixed (some nominal, categorical, interval or ratio etc.)

2.2 Goodness of Fit Test

Goodness of fit in a statistical model described how well the model fits a set of data. (Liu et al 2016).

The Pearson chi-square for testing goodness of fit is given as

$$\mathcal{X}^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] \quad (2)$$

Where $\mathcal{X}^2$ = Chi-square goodness of fit test; O_i = Observed frequency; E_i = Expected frequency; n = The sample size

2.3 Information Criterion Tests

According to Hilbe, (2009), Information Criterion tests are single statistics used by researchers and analysts to compare best fits models.

The Akaike information criterion (AIC) tests named after the Japanese statistician Hirotsugu Akaike (1927-2009), is perhaps the most well-known and well used information statistics in current research.

Traditionally AIC statistics can be mathematically defined as:

$$\text{AIC}=\frac{-2\mathcal{L}+2k}{n} \quad or \quad \frac{-2(\mathcal{L}-k)}{n} \quad (3)$$

Where $\mathcal{L}$ is the model log-likelihood, k is the number of parameter estimates in the model and n is the number of observations in the model.

The Schwarz Bayesian information criterion (BIC) is known to be the most used BIC test found in literature according to sources. Developed by Gideon Schwarz in 1978, its value differs little from the AIC statistic. Though it is the most preferred information criterion for most statisticians, the AIC however appears to be more popular.

BIC can be defined mathematically as:

$$BIC=2\mathcal{L} + kln(n) \quad (4)$$

Where $\mathcal{L}$ is the model log-likelihood, k is the number of parameter estimates and n is the number of observation in the model.

3 Data

The data consisted of 1606 registered TB patients treated between January 2015 to December 2020 at the Federal Medical Center, Keffi. Out of the 1606 patients examined, 693 representing 43.2% were females, while 913 representing 56.8%.

4 Results and Discussion

Table 1 Dynamics of HIV

hiv	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0027812	.0035282	0.79	0.431	-.004134	.0096965
sex	-.1951566	.1133227	-1.72	0.085	-.4172649	.0269517
TB	-.4737924	.1359787	-3.48	0.000	-.7403056	-.2072791
_cons	-.5900321	.1897118	-3.11	0.002	-.9618604	-.2182039

From the result presented in Table 1 above, the model indicated that TB have a statistically significant relationship with HIV (p-value=0.000<0.05). Also, gender showed a non-statistically significant relationship with HIV (p-value=0.085>0.05) as seen from the table above. In addition, age is seen to be non-statistically significant to HIV (p-value=0.431>0.05) indicating an insignificant relationship with HIV.

Table 2 Odd Ratio of the dynamics of HIV

hiv	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	1.002785	.0035381	0.79	0.431	.9958745	1.009744
sex	.8227058	.0932312	-1.72	0.085	.6588463	1.027318
TB	.6226365	.0846653	-3.48	0.000	.4769681	.8127928
_cons	.5543095	.105159	-3.11	0.002	.3821812	.8039615

Table 2 presents result for the odds ratio on HIV. The model estimates that older TB patients in age are 0.2% more likely to develop HIV infection than younger TB patients. Also, as seen in the regression analysis output table above, male gender TB patient are less likely in developing HIV infection than the female gender. Furthermore, in our model, we noticed that patients with pulmonary TB cases have a 62% greater odds of developing HIV infection compared to does with extra-pulmonary TB.

Table 3 Logistic Model Goodness of Fit test for HIV model

Logistic model for hiv, goodness-of-fit test
number of observations = 1606
number of covariate patterns = 276
Pearson chi2(272) = 313.68
Prob > chi2 0.0417

The result in Table 3 above, it shows that the model does not fit very well, as the significant level is p (0.0417 < 0.05). Therefore, the goodness-of-fit test using the Pearson chi square test shows that the models for HIV does not follow a logistic distribution and model.

Table 4 Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	1,606	-942.9953	-934.9337	4	1877.867	1899.393

Note: N=Obs used in calculating BIC; see [R] BIC note.

Both the AIC and BIC are metrics that are used to evaluate models in order to get the best fit model. They can also be referred to as selection method for selecting the best fit model. According to the table 4 above, the value 1877.867 and 1899.393 represent the models with the lowest value for both AIC and BIC and hence represent the best fit model for HIV model.

Table 5 Classification Test for the HIV model

Classified	True D	True ~D	Total
+	0	0	0
-	440	1166	1606
Total	440	1166	1606

Classified + if predicted Pr(D) >= .5
True D defined as hiv != 0

Sensitivity	Pr(+| D)	0.00%
Specificity	Pr(-|~D)	100.00%
Positive predictive value	Pr(D| +)	.%
Negative predictive value	Pr(~D| -)	72.60%
False + rate for true ~D	Pr(+|~D)	0.00%
False - rate for true D	Pr(-| D)	100.00%
False + rate for classified +	Pr(~D| +)	.%
False - rate for classified -	Pr(D| -)	27.40%
Correctly classified		72.60%

Table 5 presented the model for the classification test on HIV, which revealed to us a 72.60% level of correct classification for HIV negative and HIV positive patients. This implies a high level of agreement with our model.

Table 6 Dynamics of TB

TB	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0171662	.003963	-4.33	0.000	-.0249336	-.0093988
sex	.085819	.1294662	0.66	0.507	-.1679302	.3395681
hiv	-.4727774	.1360316	-3.48	0.001	-.7393943	-.2061604
_cons	2.168064	.1815489	11.94	0.000	1.812235	2.523894

Table 6 present the Logistic Regression analysis coefficient for TB. The model shows that HIV has a strong statistically significant relationship with TB (p-value=0.001<0.05). In addition, Age can also be seen to have a strong statistical influence on TB (p-value=0.000<0.05). Finally, from the model gender is statistically insignificant to TB (p-value=0.507>0.05) in this study.

Table 7 Odd Ratio for TB Model

TB	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9829803	.0038956	-4.33	0.000	.9753746	.9906452
sex	1.089609	.1410676	0.66	0.507	.8454129	1.404341
hiv	.6232688	.0847842	-3.48	0.001	.477403	.8137025
_cons	8.741348	1.586982	11.94	0.000	6.12412	12.47708

Table 7 above presents the odds ratio on TB model. The model reveals to us that older HIV patients in age are 98% more likely to develop TB as compared to younger HIV patients in age. Also, from the result of the analysis male gender have 8% more likely to developing TB than the female gender. And lastly from our model, HIV positive patients have 62% less likely in developing TB than the HIV negative patients.

Table 8 Logistic Model Goodness of Fit test for TB

Logistic model for TB, goodness-of-fit test
number of observations = 1606
number of covariate patterns = 280
Pearson chi2(276) = 289.96
Prob > chi2 = 0.2700

The result from the analysis in Table 8, the goodness of fit test suggests that the TB model is well fitted (p-value=0.2700>0.05). Hence, suggest that the goodness of fit test for TB follow a logistic distribution and model.

Table 9 Classification Test of Logistics model for TB

Classified	True D	True ~D	Total
+	1298	308	1606
-	0	0	0
Total	1298	308	1606

Classified + if predicted Pr(D) >= .5
True D defined as TB != 0

Sensitivity	Pr(+| D)	100.00%

```
Specificity                     Pr( -|~D)    0.00%
Positive predictive value       Pr( D| +)   80.82%
Negative predictive value       Pr(~D| -)       .%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)  100.00%
False - rate for true D         Pr( -| D)    0.00%
False + rate for classified +   Pr(~D| +)   19.18%
False - rate for classified -   Pr( D| -)       .%
--------------------------------------------------
Correctly classified                        80.82%
--------------------------------------------------
```

Table 10 shows the model for the classification test for TB which revealed to us about 80.82% level of correct classification of TB. This simply implies a high level of agreement for our model.

5 Conclusion

This study investigated the dynamics of TB and HIV among TB patients in Federal Medical Center Keffi, Nigeria using Logistic regression analysis. To achieve the aim of this study, dataset on recorded patients with Tuberculosis cases were collected from the DOT center, Federal Medical Center keffi, Nigeria. The data consisted of 1606 registered TB cases recorded between January 2015 to December 2020. Out of the 1606 TB cases examined, 693 representing 43.2% were females, while 913 representing 56.8% were male gender. The major finding revealed that HIV and TB affects each other in a bi-directional way**.** This is similar to the work of Adenomon and Akinyemi (2020).

References

Adenomon, M. O. & Akinyemi, G. S. (2020): Statistical Analysis of Tuberculosis and HIV Cases in West Africa Using Panel Poisson and Negative Binomial Regression Models. 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), IEEE 10.1109/ICMCECS47690.2020.240843. Ieeexplore.ieee.org/abstract/document/9077638. Pp. 1-6.

Awodiji, O. A., Adenomon, M. O., Ololo, K. C. and Ajayi, I. G. (2023): The Economic Impact of the COVID-19 lockdown on familylife and children's Eating Habits. EUREKA: Social and Humanities,4, 31-41

Hilbe, JM, (2009): Practical Guide to Logistic Regression

Logan, M. (2010): Biostatistical Design and Analysis using R. A practical guide. UK: John Wiley & Sons LTD

Ngowi, B.J., Mfinanga, S.G., Bruun, J.N. and Morkve, O. (2008) Pulmonary Tuberculosis among People Living with HIV/AIDS Attending Care and Treatment in Rural Northern Tanzania. BMC Public Health, 8, 341. http://dx.doi.org/10.1186/1471-2458-8-341.

Njepuome, N. and Odume, B. (2009): The Impact of HIV Syndromes on the treatment of TB cases in Gombe State, Nigeria. Mera African Journal of Respiratory Medicine, 16-20.

World Health Organization (2018) Global Tuberculosis Report, 2018. WHO, Geneva, Switzerland.

World Health Organization (2020) Global Tuberculosis Report, 2020. WHO, Geneva, Switzerland.

Zar, J. H. (2014): Biostatistical Analysis (5th ed.). UK: Pearson Education Limited