E-ISSN: 1116 - 3089 P-ISSN: 3026-9784
MPRI-JOMDR | INTERNATIONAL JOURNAL OF ORGANIC & MEDICINAL DEV. RESEARCH, VOL. 03 NO. 3 JANUARY, 2024

LOGISTIC REGRESSION ANALYSIS ON FACTORS CONRIBUTING TO THE ACQUISITION OF HIV INFECTION AMONG TUBERCULOSIS PATIENTS

[1]FASASI, S.K.; [1]OSUNTOKI, N.B.; & [2]SALAU, M.A.

[1]Department of Mathematics and Statistics, School of Science, The Oke – Ogun Polytechnic, Saki, Oyo State. [2]Department of Agricultural Extension & Management Technology, The Oke – Ogun Polytechnic, Saki, Oyo State.
shafykay2012@gmail.com

Abstract

The study aimed at examining the factors that contribute to the acquisition of HIV infection among TB patients. Results from the analyes show that out of 736 TB patients screened for HIV infection within the study period, the percentage of female TB patients that are HIV positive are more than male TB patients. The findings also reveal that HIV positivity rate was found to be highest in TB patients in the middle age-group (25-49), married TB patients, self-employed TB patients and TB patients with no education among others. A chi-square test of independence was carried out to test if HIV status of TB patients is dependent of age, gender, marital status, occupation and level of education, which shows that it is dependent of age, gender, marital status, occupation and educational level. A logistic regression analysis was also carried out to formulate a model that

Keywords: Contribution, HIV Infection, Tuberculosis, Patients, Logistic Regression.

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

can be used to predict the HIV status of TB patient using age, gender, marital status, occupation and educational level. The model shows that some of the categories of the predictor variables does not have any link to HIV acquisition, the model also shows that TB patients that are male, age (25-49), student, secondary education, higher education have some link to HIV acquisition.

INTRODUCTION

Human immunodeficiency virus (HIV) and Mycobacterium tuberculosis (TB) infections remain the leading causes of morbidity and mortality from infectious diseases in both children and adults in sub-Saharan Africa (Corbett E., 2016). Most recent global estimates of incident cases of HIV and active TB infections stood at 2.5 million and 9 million respectively with 13% of TB cases co-infected with HIV (WHO global report, 2019). The mechanisms of susceptibility of tuberculosis patients to co-infection with HIV remain largely unknown and promoted by a complex set of socio-behavioral and environmental factors (Otu A., 2013). The overall aim of this study was to determine the factors contributing to the acquisition of HIV infection among tuberculosis patients.

Nigeria has recorded an overall steady decline in the proportion of people infected with HIV from a peak of 5.8% in 2001 to 3.2% in 2013.the country remains the second worst HIV affected country in the world with a total estimated burden of 3.4 million people living with HIV/AIDS (UNAIDS 2018). There has been a rapidly growing trend in the epidemiologic synergy forged between the two diseases particularly in sub-Saharan Africa as one potentiates the other and complicate other unrelated infectious diseases among vulnerable groups. While the literature is sparse with regard to attempt to define the prevalence magnitude of the rising co-epidemic and its socio-economic dimension in south-west Nigeria, a few studies have signaled the potential of the two diseases to erode past public health gains a well a slowing down effectiveness of current control efforts in Nigeria (Otu, A., 2013).

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

Furthermore, TB is known to be preventable and curable; however, it remains the most common cause of AIDS- related deaths in Nigeria and sub-Saharan Africa. Conversely, among individuals with latent TB within a population, HIV remains the most significant cause of reactivation of tuberculosis, and rapid progression to death if not diagnosed early or effectively treated. However, being a disease that thrives in poor living conditions, TB ensures that effective control can be attained by proper implementation of sanitary and hygienic practices which include avoidance of overcrowding (especially in households, schools, correctional and health facilities) and the minimization of air pollution (Hargreaves J., 2011). Evidence of high burden TB reactivation has also been found among immigrants in countries with low incidence of tuberculosis as reported by Lillebaek and her colleagues (Lillebaek T. et al, 2015).

In Nigeria, few studies have assessed presentation and treatment outcome of HIV and TB co-infection both in children and adult TB patient populations. A five-year retrospective assessment conducted by Daniel and his colleagues in Ogun state reported a 10.5% period prevalence of HIV among 76 children with TB (Daniel, O.J., 2007). In a separate study conducted by Umeh and others among patients who presented with respiratory symptoms in a referral chest clinic, the risk of co-infection was about three times higher among HIV positives than TB positives (Umeh E et al, 2017). Certain socio-cultural factors have been shown to exacerbate co-infection in Nigeria which include gender inequality and sexual violence entrenched in diverse Afro-cultural norms particularly among rural and urban dwellers in sub-Saharan Africa (Pennap G. et al, 2010).

Aim and Objectives

The aim of this study is to examine the factors contributing to the acquisition of HIV infection among TB patients and the specific objectives are to:

(i) Examine the percentage of TB patients that are co-infected with HIV using frequency distribution table;

(ii) Examine if the HIV status of TB patient is dependent on the factors, namely- age, gender, marital status, occupation and educational level;

(iii) Determine the relationship between the HIV status of the B patients with respect to their age, gender, marital status, occupation and education level

The scope of the study in this research work was limited to in-patients with tuberculosis at the General Hospital, Saki, Oyo State

LITERATURE REVIEW

Logistic regression is a technique for making predictions when the dependent variable is dichotomous in nature and the predictor variable(s) is/are continuous and/or discrete (Park RS et al, 2021). It assumes that the relationship between the independent variable(s) and the dependent variable is logarithmic. It allows one to predict a discrete outcome such as group membership from a set of variables that may be continuous, discrete, dichotomous or a mix (Vetter T.R. & Schober P. 2018). It is intended to predict probabilities, but these are between '0' and '1'. Values between '0' and '1' would be difficult to predict by a linear combination of predictors, because the range usually is not limited for the outcome variable. It is also more flexible than discriminant analysis or regular regression analysis, since unlike discriminant analysis, logistic regression does not assume continuous, normally distributed predictors and regression analysis, logistic regression does not run into the problem of predicting negative probabilities for group membership (Raymond S. 2021).

This work is an application of logistic regression to factors contributing to the acquisition of HIV infection among tuberculosis patients. Since the outcome variable in the study are group in HIV positive and negative, and the predictor variables are all categorical then logistic regression is the best method to use. It's also known as logit model. It computes the probability of the selected response as a function of the values of the predictor variables. If a predictor variable is categorized into two values, then one of the values is assigned the value '1' and the other is assigned the value '0'. In logistic regression, the dependent(outcome) variable is a logit, which is the natural logarithm of the odds (Raymond S. 2021).

THE LOGISTIC REGRESSION EQUATION

Logistic regression can be binomial, ordinal or multinomial. Binomial or binary logistic regression deals with situations in which he observed outcome for a dependent variable can have only two possible types '0' and '1' (which may represent, for example, success or failure, absence or presence, yes or no etc.). In binary logistic regression, the outcome is usually coded as '0' or '1', as this leads to the most straightforward interpretation (Homer, David W; Lemeshow, Stanley, 2000). If a particular observed outcome for the dependent variable is noteworthy possible outcome (referred to as 'success') it is usually coded as '1' and the contrary outcome (referred to as 'failure') as '0'. In regression analysis, logistic regression is estimating the parameters of a logistic model; it is a form of binomial regression. Linear regression may make nonsensical predictions for a binary dependent variable. What is needed is a way to convert a binary variable into a continuous one that can take on any real value (negative or positive). To do that, binomial logistic regression first calculate the odds of an event happening for different levels of each independent variable, and then takes its logarithm to create a continuous criterion as a transformed version of the dependent variable. The logarithm of the odds is the logit of the probability, the logit being defined as follows:

$$\text{Logit } (p(y)) = In\frac{p(y)}{1-p(y)} \qquad \text{for } 0<p(y)<1$$

Where,

p(y) is the probability that the dependent variable equals a case.

Although the dependent variable in logistic regression is Bernoulli, the logit is on an unrestricted scale (Homer, David W.; Lemeshow, Stanley, 2000). The logit function is the link function in this kind of generalized linear model, i.e.it shows the relationship between the usual regression equation, which is a straight line formula and the logistic regression equation. The form of the logistic equation is:

$$\text{Logit } E(y) = \alpha+\beta x$$
$$= \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots$$

Where Y is the Bernoulli-distributed response variable and x is the predictor variable.

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

The logit of the probability of success is then fitted to the predictors. The predicted value of the logit is converted back into predicted odds via the inverse of the natural logarithm, namely the exponential function. Thus, although the observed dependent variable in binary logistic regression is a zero-or-one variable, the logistic regression estimate has the odds as a continuous variable, that the dependent is a success (a case) (Penduzi 2015). The probability of success can be calculated using the inverse of the logistic function. i.e. the inverse of the logistic function, g, the logit (log odds) is defined as:

$$g(p(y)) = \text{logit}(p(y)) = \ln \frac{p(y)}{1-p(y)} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots$$

where

$$p(y) = \frac{e^{(\alpha+\beta 1x1+\beta 2x2+\cdots)}}{1+e^{(\alpha+\beta 1x1+\beta 2x2+\cdots)}}$$

and

$$1 - p(y) = 1 - \frac{e^{(\alpha+\beta 1x1+\beta 2x2+\cdots)}}{1+e^{(\alpha+\beta 1x1+\beta 2x2+\cdots)}}$$

Equivalently, after exponentiating both sides:

$$\frac{p(y)}{1-p(y)} = e^{(\alpha + \beta 1x1 + \beta 2x2 + \beta 3x3 + \dots)}$$

Where

g is the logit function

In denotes the natural logarithm

P(y) is the probability that the dependent variable equals a case, given some linear combination of the predictors.

α is the intercept from the linear regression equation (the value of the criterion when the predictor(s) equal zero)

β is the regression coefficient of the predictors

base e denotes the exponential function

Table 1: FREQUENCY (PERCENTAGE) OF SOCIO-DEMOGRAPHIC DISTRIBUTION OF STUDY POPULATION (TB patients) IN RELATION TO HIV STATUS

Factors/levels	Total TB patients enrolled	HIV positive Frequency (%)	HIV negative Frequency (%)
Gender			
Male	410	95 (23.17)	315 (76.83)

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

Female	326	106 (32.52)	22 (67.48)
Age			
0-24	170	22 (12.94)	148 (87.06)
25-49	362	122 (33.7)	240 (66.3)
≥50	204	57 (27.94)	147 (72.06)
Marital Status			
Married	442	153 (34.62)	289 (65.38)
Single	251	36 (14.34)	215 (85.66)
Widowed/Divorced	43	12 (27.91)	31 (72.09)
Occupation			
Civil Servants	59	11 (18.64)	48 (81.36)
Unemployed	21	3 (14.29)	18 (85.71)
Student	191	17 (8.9)	174 (91.1)
Others	465	170 (36.56)	295 (63.44)
Educational level			
None	35	15(42.86)	20 (57.14)
Primary	193	61(31.61)	132 (68.39)
Secondary	388	111(28.61)	277 (71.39)
Higher	120	14 (11.67)	106 (88.33)

Table 2: Contingency table showing the observed and the expected counts, Chi-square value and p-value of HIV status of TB patients and the independent variables

Factors	HIV status of TB patients				Chi-Square value	P-value
	Observed Counts		Expected Counts			
	Positive	Negative	Positive	Negative		
Gender						
Male	95	315	112.0	298.0		0.005
Female	106	220	89.0	237.0	7.988	
Age						
0-24	22	148	46.4	123.6		
25-49	122	240	98.9	263.1		<0.0001
≥50	147	57	55.7	55.7	25.171	
Marital status						
Married	153	289	120.7	321.3		
Single	36	215	68.5	182.5		<0.0001
Widowed/Divorced	12	31	11.7	31.3	33.153	
Occupation						

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

Civil Servants	11	48	16.1	42.9	56.673	<0.0001
Unemployed	3	18	5.7	15.3		
Student	17	174	52.2	138.8		
Others	170	295	127.0	338.0		
Educational level					21.178	<0.0001
None	15	20	19.6	25.4		
Primary	61	132	52.7	140.3		
Secondary	111	277	106.0	282.0		
Higher	14	106	32.8	87.2		

Decision: Since the p-values for all the factors are less than α (0.05), we reject H_0

Conclusion: we conclude that the HIV status of the TB patients is dependent of Age, Gender, Marital status, Occupation and level of education.

Table 3: Classification Table of HIV status of the TB patients

Classification Table					
	Observed		Predicted		
			HIV status of the TB patients		Percentage Correct
			Negative	Positive	
	HIV status of the TB patients	Negative	535	0	100.0
		Positive	201	0	.0
	Overall Percentage				72.7

Table 3 presents the results with only the constant included without any Predictor Variables (Age of the patients, Gender, Marital status, Occupation and Educational level) in the equation. The prediction of this Intercept-only model is made purely on whichever category occurred most in the dataset. In this research the model always guesses "Negative" because more of the TB patients are HIV negative. The Overall Percentage row tells that this approach to prediction is correct 72.6% of the time.

Table 4: Classification Table of HIV status of the TB patients with respect to the Predictor Variables

Classification Table			
Observed	Predicted		
	HIV status of the TB patients		Total
	Negative	Positive	

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

HIV status of the TB patients	Negative	520	15	97.2
	Positive	178	23	11.4
Overall Percentage				73.8

The Classification Table correctly classify 23/201=**11.4**% of the TB patients to the **HIV positive category** (predicted event) observed. This is known as the **sensitivity** of prediction (the percentage of occurrences correctly predicted). It also correctly classifies 520/535=**97.2**% of the TB patients to the HIV negative category (non-predicted event) observed. This is known as the specificity of prediction (percentage of non-occurrences correctly predicted). Overall prediction were correct 543 out of 736 TB patients, for an overall success rate of 73.8% compare to probability of a correct prediction 72.7%, therefore there is a slight improvement in the prediction.

Table 5: Omnibus test of model coefficients

Omnibus Test of Model Coefficients				
		Chi-Square	Df	Sig.
	Step	99.050	11	<0.001
	Block	99.050	11	<0.001
	Model	99.050	11	<0.001

Table above indicates that the predictor variables entered the model simultaneously. The method used is Enter.

H_0: The predictor variables does not have significant effect in the model

H_1: The predictor variables have significant effect in the model

Omnibus Test of Model Coefficients gives a chi-square value of 99.05 on 11 df: significant beyond 0.001 Thus, the null hypothesis is rejected.

Table 6: Model Summary

Model Summary			
	-2 Log likelihood	Cox & Snell R squared	Nagelkerke R squared
	764.007	0.126	0.182

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

Under model summary, we see that the -2Loglikelihood statistic is 764.007. This statistic test whether a set of predictor variables improves prediction. Adding the Predictor variables (Age, Gender, Occupation, Marital Status and Educational level) reduced the -2LL by 99.05, therefore, we conclude that at least one of the covariates is significant in explaining the variation in HIV status of TB patients. Pseudo R squared: Nagelkerke R squared is preferred, the model accounts for 18% of variance of HIV status (Dependent Variable).

Table 7: Variables in the Equation

Parameter	Estimate B	Standard Error	Wald	df	P-value	Exp(B) Odd Ratio	95% C.I for Exp(B) Lower	Upper
Intercept	0.278	0.785	0.125	1	0.723	1.320		
Gender								
Male	-0.594	0.188	10.003	1	0.002	0.552	0.382	0.798
Age-Group								
Age 25-49	0.112	0.436	0.066	1	0.797	1.060	0.881	2.100
Age≥50	-0.967	0.485	3.966	1	0.046	0.380	0.147	0.985
Marital Status								
Married	0.625	0.407	2.357	1	0.125	1.869	0.841	4.152
Single	0.256	0.513	0.249	1	0.618	1.291	0.473	3.528
Occupation								
Unemployed	-1.211	0.829	2.133	1	0.144	0.698	0.359	1.513
Student	-1.196	0.579	4.274	1	0.039	0.302	0.097	0.940
Others	0.637	0.427	2.233	1	0.135	1.891	0.820	4.364
Educational Level								
Primary	-0.735	0.445	2.725	1	0.099	0.479	0.200	1.148
Secondary	-1.603	0.462	12.029	1	0.001	0.201	0.081	0.498
Higher	-1.783	0.596	8.959	1	0.003	0.1680	0.052	0.540

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

From Table 7, Marital Status, Age 25-49, Unemployed, Others and Primary Education are not significant in explaining the variation in HIV status of the TB patients but all other categories of the factors considered are significant, therefore, Interpretation of the significant covariates are given.

The model was formulated as follows: For the HIV status of the TB patients, let Y_i be the binary outcome for HIV status (positive/negative) for individual TB patients, $Y_i \sim \boldsymbol{Bernoulli}$ (HIV positive).

$$In\left(\frac{Y=HIV\ positive}{1-Y=HIV\ negative}\right) = \beta_0 + \beta_1\ Male + \beta_2\ Age{\geq}50 + \beta_3\ Student + \beta_4\ Secondary\ Education + \beta_5\ Higher\ Education.$$

Thus using the estimates of the parameters in table 3.7, the binary logistic regression model is given as:

$$In\left(\frac{Y=HIV\ positive}{1-Y=HIV\ negative}\right) = 0.278 - 0.594\ Male - 0.967\ Age{\geq}50 - 1.196\ Student - 1.603\ Secondary\ Education - 1.783 Higher\ Education$$

INTERPRETATION OF ODDS RATIOS

Gender: Male TB patients are 44.8% less likely to be HIV positive than Female TB patients (reference category).

Age: TB patients between Age (25-49) are 1.06 times more likely to be HIV positive compared to TB patients between Age (0-24) (reference category) while TB patients between Age≥50 are 62% less likely to be HIV positive compared to TB patients between Age (0-24).

Marital Status: Married and Single TB patients are 1.869 and 1.291 times respectively more likely to be HIV positive compared to TB patients that are Widowed/Divorced (reference category).

Occupation: Self-employed(Others) TB patients are 1.891 times more likely to be HIV positive compared to TB patients that are Civil Servants while Unemployed TB patients are 30.2% less likely to be HIV positive compared to TB patients that are Civil Servants. TB patients that are students are 69.8% less likely to be HIV positive compared to Civil servant TB patients.

Educational level: The odds of HIV infection decreases as the level of education increases. TB patients with primary education, secondary education and higher

education are 52.1%, 79.9% and 83.2% respectively less likely to be HIV positive compared to TB patients with no education.

CONCLUSION

Based on the study and the results obtained from the analysis, it is concluded that HIV status of TB patients is affected by age, gender, marital status, occupation and educational level. Also, middle-aged female TB patients with low or no education are the most at risk of acquiring HIV infection.

RECOMMENDATION

The analysis revealed that in the study population, middle-aged female TB patients with low or no education are the most at acquiring HIV infection. Therefore, HIV prevention effort and integrated response targeted at this sub-population of TB patients in high-risk communities across the country will go a long way in significantly stemming the rising tide of HIV-TB co-infection among individuals with underlying TB infection.

REFERENCES

Agresti C., Kwaru A.H., Magashi M.A., Nwokedi E.E. (2006), Prevalence of human immunodeficiency virus among tuberculosis patients in Kano, Nigeria, Highl. Med. Res. J. 6 (1-2).

Amuta, E., Mary, T., Ishaku, A. (2013), A retrospective study on the epidemiological trend of human immunodeficiency virus (HIV) and pulmonary tuberculosis (PTB) coinfection in Nasarawa State, Nigeria, J. Nat. Sci. Res. 3 (7) 45-51.

Bagley, S. C., H. White, and Golomb, B. A. (2001), Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. Journal of Clinical Epidemiology 54(10) 979-985.

Chuks, O., Anyabolu, A., Martin, I., Stephen, K. (2012), Prevalence of HIV infection in pulmonary tuberculosis suspects; assessing the Nnamdi Azikiwe University Teaching Hospital, Nnewi, Nigeria, Adv. Life Sci. Technol. (14) 87-90.

Concarto J., Feinstein A. R., and Holford T. R. (2003), The Risk of determining risk with Multivariable models. Annals of internal Medicine 118 (3) 201-210.

Hargreaves J., Boccia D., Evans C., Adato M., Petticrew M., Porter J. (2019), The social determinants of tuberculosis: from evidence to action, Am.J. Public Health 101(4).

Kwan C., Ernst J. (2018)., HIV and tuberculosis: a deadly human syndemic, Clinical Microbiology Rev. 24 (2) 351-376.

Lillebaek, T., Anderson, A., Dirksen, A., Smith, E. (2018), Persistent high incidence of tuberculosis in immigrants in a low-incidence country, Emerg. Infect. Dis. 8 (7) 679-684.

Ofoegbu, O.S., Odume, B.B., (2015), Treatment outcome of tuberculosis patients at National Hospital Abuja Nigeria: a five year retrospective study, S. Afr. Fam. Pract. 57 (1) 50-56.

Otu. A. (2018), A review of the national tuberculosis and leprosy control programme (NTBLCP) of Nigeria: challenges and prospects, Ann. Trop. Med. Public Health 6(5) 491-500.

Penduzzi K., Daftary A., Padayatchi N. (2015), Social constraints to TB/HIV healthcare: accounts from coinfected patients in South Africa, AIDS Care 24 (12) 1480-1486.

Raymond S., Daniel O.J., Ogunfowora O.B., Oladapo O.T. (2021), HIV sero-prevalence among children diagnosed with TB in Nigeria Trop.Doct. 37 (4) 268-269.

Samantha B., Bird, G. L., Kuijpers, M., Zimmerman, R.A., Jarvik, G. P., Wernovsky, G., Clancy, R. R., Licht, D. J., Gaynor D. J., and Nataraj, M.(2013), Prediction of periventricular leukomalacia. Part 1: Selection of hemodynamic features using logistic regression and decision tree algorithms. Artificial Intelligence in Medicine 46(3): 201-215.

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]

Umeh E., Ishaleku D., Iheukwumere C., HIV/tuberculosis coinfection among patients attending referral chest clinic in Nasarawa state, Nigeria, J. Appl. Sci. 7 (6) (2017) 933-935.

UNAIDS, UNAIDS Report on the Global AIDS Epidemic, 2018 Available at: http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/epidemiology/2018/gr2018/UNAIDS_Global_Report_2018_en.pdf.

Veteran,P.,Stephenson, R.,Winter, A., Elfstrom, M. (2009), Community environments shaping transactional sex among sexually active men in Malawi, Nigeria and Tanzania, AIDS Care 25(6) 784-792.

Yusuf, M., Azeez-Akande, Yusha'u, M. (2013), HIV sero-prevalence among adult with newly diagnosed pulmonary tuberculosis in Kano, Nigeria, J. Med. Trop. 15(2) 140-143.

WHO Global Tuberculosis Report, 2019 Available at: http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf

EDITOR-IN-CHIEF
Prof. A.S. Hossein [Saskatchewan]