LOGISTIC MODELING OF TUBERCULOSIS TREATMENT OUTCOMES FROM NON-DISJOINT EVENTS

Agada, P. O.,[1*] Egahi, M.,[2] Igbabul, S. A.[3]

[, 1] Department of Mathematics/Statistics/Computer Science, University of Agriculture Makurdi, Benue State, Nigeria*

[2,3] Department of Mathematics and Computer Science, Benue State University, Makurdi, Benue State, Nigeria

Email: gadexx@yahoo.com gadexx@yahoo.com,[2] musaegahi@gmail.com,[3] igbabulsa@gmail.com

Abstract

In Multinomial Logistic Regression, result interpretation can be difficult if an event that gives rise to at least one of the treatment outcomes is a merger of outcomes of two other non-disjoint events. This is obvious as the non-disjoint nature of the underlying events would not yield distinct and independent outcomes of the dependent variable. This work proposes a way of handling challenges of this nature at the primary level of data collection and when the study data is secondary with dependent variable outcomes overlaps. This work demonstrates how to remove these outcomes overlaps for clearer result interpretation using a dataset from a published research. The aforementioned research is on the evaluation of tuberculosis treatment outcome of TB/HIV Co- Infection. Our study made two assertions for removing outcome overlaps leading to the fitting of two Binary Logistic Regression Models one for each. Major results of the study include the fact that, if all successfully treated TB patients are cured or not, HIV status remains a predictor of TB treatment outcome with decrease in the odds for patients who test positive to HIV relative to those who test negative. Additional result shows that if all the successfully treated TB patients are cured, then, sputum test result for TB at baseline and the sex of patient become immaterial in predicting TB treatment outcomes. Rather, the fact that patients are on treatment support and on ART become significant predictors, with odd ratios of 1.644 and 0.759 respectively. Furthermore, if all the successfully treated TB patients resisted treatment (not cured), the predictors of TB treatment outcome excludes the fact that patients are on treatment support and on ART but include their sex and sputum test result at baseline, with odd ratios of 0.645 and 54.938 respectively.

1.0 Introduction

The Multinomial Logistic Regression Model has been employed in modeling the treatment outcomes of diseases most especially when there are more than two (2) distinct and independent outcomes. However, interpretation of model results can be difficult if an event that gives rise to one of the treatment outcomes is a merger of the outcomes of two other non-disjoint events. This is obvious as the non-disjoint nature of the underlying events would not yield distinct and independent outcomes of the dependent variable.

A way of avoiding challenges of this nature is by clearly defining disjoint events of the dependent variable that will yield distinct and independent outcomes. This we advice should be done at the primary level of data collection on the subjects of the study as it will no doubt, ensure dependent variable outcome non-overlap.

Moreover, if the data is a secondary data with dependent variable outcomes overlaps, then the overlaps need be resolved if model results would be meaningful. This work proposes a way for this resolve as follows; firstly, identify and separate the dependent variable with the problem of overlap, Secondly, merge the remaining non overlapping outcomes broadly into binary outcomes (success and failure).

Furthermore, assert that; the event that gave rise to the dependent variable outcomes overlaps in question, is from the merger of two events which are disjoint or compliments of themselves (i.e no overlaps). These yield two binary outcomes of success and failure respectively for the event and its compliment. These outcomes are finally merged with the aforementioned binary outcomes of success and failure in step 2 above to form distinct and independent outcomes of the dependent variable. The first Binary Logistic Regression Model relating these binary outcomes to a set of predictor variables or risk factors are obtained via this assertion.

Another relevant assertion that can be made is that, event that gave rise to the dependent variable outcomes overlaps in question, is from the merger of two same events. The outcomes of this merger, forms those of either the success or failure event of the binary response variable. As in the earlier assertion, these outcomes are finally merged with the aforementioned binary outcomes of success and failure in step 2 to form distinct and independent outcomes of the dependent variable. The second Binary Logistic Regression Model relating the binary outcomes to a set of predictor variables or risk factors are obtained via this second assertion.

We envisage that this will help eliminate the effect of the non-disjoint events as it affects the interpretation of model results. On the contrary, one may suggest the use of the multinomial regression model after the problem of dependent variable overlap is resolved. This option though, depends on the interest of the researcher; it also depends on having adequate sample data size. This is because, the non-merging of the dependent variable outcomes into broad binary outcomes of success and failure as earlier proposed, might result into poor model fit due to inadequate data (Peduzzi, Concato, Kemper, Holford and Feinstein, 1996).

In order to demonstrate our proposal, the same dataset on Tuberculosis (TB) treatment outcomes and predictor variables used by Hassan, Olukolade, Ogbuji, Onyemocho, Okwuonye, Igbabul, Okechukwu, Kusimo, Osho, Osinowo and Ladipo (2006) was employed in this work. The authors related TB treatment outcomes (Cured, Failed, Successful treatment, Defaulted, Transferred and Death) with some predictor variables or risk factors using a Multinomial Logistic Regression Model. The predictor variables or risk factors include; Age $(< or \geq 45 yrs)$, Sex (male or female), Treatment support (yes or no), HIV status (yes or no), Baseline sputum result (positive or negative), On Anti Retroviral Therapy (ART) (positive or negative) and and On Cotrimoxazole Prophylaxis Therapy (CPT) (yes or no).

According to the authors, the dependent variable outcome, tagged successful treatment is a merger of patients who were cured and those who completed treatment. We argue that the patients who completed treatment may be cured or not cured resulting to the problem of outcome overlap. Hence, we advocate for the use of distinct and independent binary treatment outcomes (good outcome (cured) and bad outcome (not cured)). To this end, we apply the two assertions made earlier and propose a Binary Logistic Regression Model relating these outcomes to the set of predictor variables or risk factors. This we believe will make better meaning than the use of response outcomes of non-disjoint events in the Multinomial Logistic Regression Modeling approach proposed by the authors.

The Binary Logistic Regression Model has been extensively and successfully used in disease treatment outcome modeling and in other areas of healthcare research. Some works include those of Mohammed and Alnory (2020) on the logistic regression analysis to determine cardiovascular disease risk factors, Coaster-Veiga, Teodoro and Nunes (2018) on unsuccessful treatment in pulmonary tuberculosis and that of Peter (2020) on the determinants of unsuccessful treatment outcomes and mortality among tuberculosis patients in Malaysia. Other works include those of Sharareh, Niakan, Mahshid and Xiao-Jun (2010), Medhin and Biadgilign (2013) and Teshome and Anagaw (2017) to mention a few.

The rest of the paper is sectioned into Methodology, Result and Discussion, Conclusion and Recommendation.

2.0 Methodology

The methodology for this work is sectioned into the Nature and source of data, Venn diagram of response variable outcomes from disjoint and non- disjoint events and the Mathematical theory of the Binary Logistic Regression Model.

2.1 Nature and source of data

The same dataset on Tuberculosis (TB) treatment outcomes and predictor variables on 2,636 subjects were used by Hassan *et al*. (2006) was employed in this work. The authors related TB treatment outcomes (Cured, Failed, Successful treatment, Defaulted, Transferred and Death) with some predictor variables or risk factors using a Multinomial Logistic Regression Model. The predictor variables or risk factors include; Sex (male or female), Treatment support (yes or no), HIV status (yes or no), Baseline sputum result (positive or negative), On Anti Retroviral Therapy (ART) (positive or negative) and On CPT drug (yes or no).

2.2 Venn diagram of response variable outcomes from disjoint and non-disjoint events

As earlier mentioned, Hassan *et al*. (2006) stated that the dependent variable outcome, tagged successful treatment is a merger of patients who were cured and those who completed treatment. We hypothesize that the patients who completed treatment may be cured or not cured. This we envisage will cause the problem of outcome overlap due to the possibility of having response variable outcomes drawn from non-disjoint events. We therefore advocate for the use of distinct and independent binary treatment outcomes tagged good outcome (cured) and bad outcome (not cured). We illustrate using Venn diagrams of response variable outcomes from disjoint and non-disjoint events.

The Venn diagram representation of the successful treatment (S) outcome being a merger of the patients who were originally confirmed cured (O) and those who completed treatment (T) (bearing in mind that those who completed treatment may be cured (C) or may not be cured (C')) is given below.

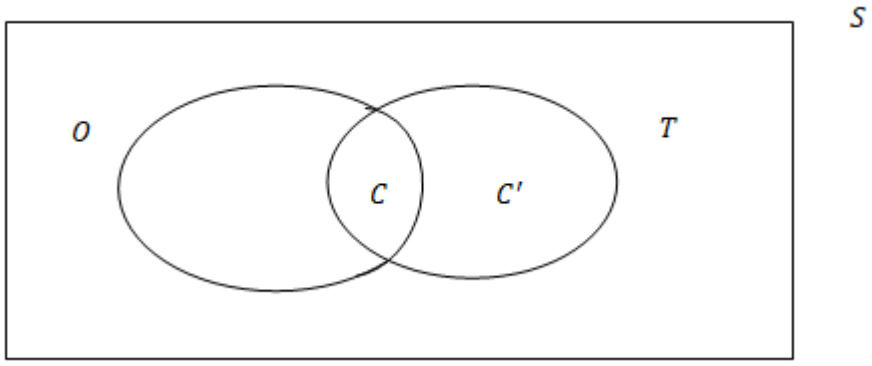

Figure 1: Response variable outcomes from non-disjoint events of cured (O) and patients who completed treatment (T)

From the Venn diagram,
$S = OT$, where OT is a concatenation of the outcomes of O and T, $T = C \cup C'$ and $C = O \cap T$
We use the term concatenation instead of union since the intersection C is repeated in the merger S.

As earlier argued, using O and S as outcomes of a binary logistic or some of the outcomes of a multinomial logistic regression model will result to difficulty in result interpretation. This is due to the outcomes overlaps as a result of the non – disjoint events that gave rise to the outcomes of the event set C. Our task is to obtain distinct and independent outcomes of the response variable in the logistic regression model via disjoint events. This can be achieved by using the two assertions earlier made in section 1.0. According to the first assertion; all patients who completed treatment (T) were not cured. This creates dis-joint events of O and T and the Venn diagram in figure 1 becomesl

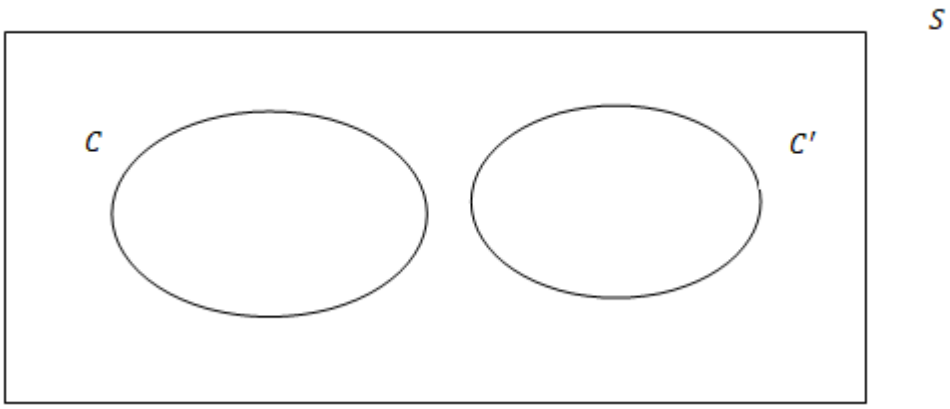

Figure 2: Response variable outcomes from disjoint events of cured and uncured patients

Where:

$S = OT$, but $T = C'$ and $O = C$, OT is as earlier defined.

The outcomes of the event that gave rise to C and C' can be merged and used as part of the dataset of the response variable in the Binary Logistic Regression Model earlier advocated for in section 1.0. The full dataset is obtained by addition of the outcome dataset from the other multinomial regression outcomes (Failed, Defaulted, Transferred and Death) of Hassan et al. (2006). This is done by adding the outcomes of the Failed, Defaulted, Transferred and Death to those of C'.

According to the second assertion; all patients who completed treatment (T) are cured. This creates a merger of two sets containing cured patients (C). The Venn diagram in figure 1 becomes;

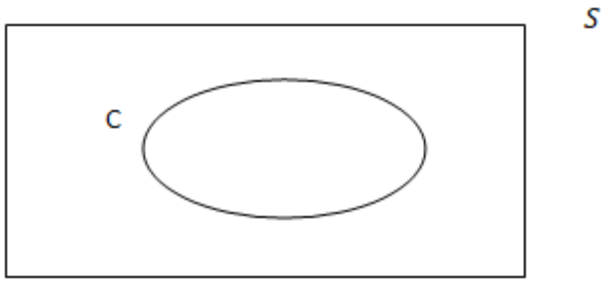

Figure 3: Response variable outcomes from the events of cured patients

The outcomes of the event that gave rise to C are merged with those of C'. C' here, refers to the outcome dataset from the other multinomial regression outcomes (Failed, Defaulted, Transferred and Death) of Hassan et al. (2006). The merger dataset forms the binary outcomes (similar to those of figure 2) of the response variable for the second Binary Logistic Regression Model earlier mentioned in section 1.0.

2.3 Mathematical theory of the Binary Logistic Regression Model

2.3.1 Model description

In Binary logistic regression modeling, models are built for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two outcomes). The main objective of binary regression modeling is to find a model that best describe the existing relationship between the dichotomous characteristic of interest and the set of independent or predictor variables. The model achieves this by generating values of model parameters for predicting a logit transformation of the probability of the presence of the characteristic of interest:

$$Logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K \quad (1)$$

It follows from equation (1) that

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_K X_K)}} \quad (2)$$

Where p is the probability of the presence of the characteristic of interest, the β_i's and the $X_i's, i = 1,2,\ldots,k$ are the regression coefficient and the independent variables respectively. In this work, p is the probability of a good treatment outcome, the independent or predictor

variables are as stated in section 2.1. The logit transformation is defined as the logged odds where;

$$Odds = \frac{p}{1-p} = \frac{probability\ of\ the\ presence\ of\ the\ characteristic}{probability\ of\ the\ absence\ of\ the\ characteristic} \quad (3)$$

and

$$Logit(p) = Ln\left(\frac{P}{1-P}\right) \quad (4)$$

Unlike the ordinary regression model that chooses model parameters that minimize the sum of squares of errors, logistic models choose parameters that maximize likelihood of observing the sample values. The Forward wald Stepwise (Likelihood Ratio) regression method is employed in the modelling process.

2.3.2 Model parameter estimation

The most commonly used method of estimating the parameters of a logistics regression model is the method of maximum like hood (ML). Generally, the sample likelihood function is defined as the joint probability function of random variables. Specifically, suppose $(x_1, x_{2,..}, x_k)$ are k independent random observations. Since y_i is the Bernoulli random variable with probability $p(x_i)$ if y_i= 1 or is $1 - p(x_i)$ if $y_i = 0$. The likelihood function for a logistic regression model is;

$$L\left(\beta_{0,}, \beta_1, \beta_2, \dots, \beta_k\right) = L = \prod_{I=1}^{n} p(x_i)^{y_i}[1 - p(x_i)]^{1-y_i} \quad (5)$$

$$\log(L) = \sum_{i=1}^{n} y_i \log p(x_i) + (1 - y_i)\log[1 - p(x_i)]$$

$$= \sum_{i=1}^{n} -\log(1) + e^{(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_KX_K)} + \sum_{i=1}^{n} y_i(\beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_KX_K) \quad (6)$$

(Mohamed and Mohamed, 2018).

To find the ML estimate, the log likelihood is differentiated with respect to each parameter and equated to zero. Since the equation is nonlinear in β, some special methods such as the Ierative Re-weighted Least or the Newton Raphson Method can be employed.

2.3.3 Model goodness of fit and adequacy checks

In order to test for the goodness of fit of the model, the -2 Log likelihood Null and Full models are employed. The Null model -2 Log likelihood is given as;

$$-2 * Ln(L_0) \quad (7)$$

Where L_0 is the likelihood of obtaining the observations if the independent variables are not included in the model (i.e have no effect on the outcome) while the Full model -2 Log likelihood is given as:

$$-2 * Ln(L) \quad (8)$$

Where L is the likelihood of obtaining the observations if the independent variables are included in the model (i.e they do have effect on the outcome).

The difference in these two yields a chi-square statistic which is a measure of how well the independent variables affect the outcome or dependent variable. If the p value for the overall model fit statistic is less thaan 0.05, then there is evidence that at least one of the independent variables contributes to the prediction of the outcome.

The Hosmer-Lemeshow test is a test of goodness of fit employed in this work. The test divides the test data into approximately 10 groups. The chi-square statistic for this test is computed by;

$$\chi^2_{HL}=\sum_{g=1}^{G}\frac{(O_g-E_g)}{E_g(1-E_g/n_g)} \quad (9)$$

with $O_g, E_g \; and \; n_g$ defined as the observed events, expected events and number of observations for the g[th] decile group and G the number of groups. The number of degree of freedom is G-2. A large value of chi-square with small p-value < 0.05 indicates poor fit while a small chi-square vale with p –value closer to 1 indicate a good logistic regression model fit.

In order to evaluate the prediction accuracy of the Binary logistic model, the classification table is employed. On this table, the observed values of the dependent variable and the predicted values at a user defined cut - off value are cross classified.

The Walds statistic tests the significance of model parameters. This helps to determine whether or not an independent variable stays in the model as it tests if the associated model parameter differs significantly from zero. The Walds statistic is computed as the regression coefficient divided by its standard error squared:

$$\left(\frac{\beta}{SE}\right)^2$$

Where $\beta = the \; associated \; regression \; parameter, SE = It'standared \; error.$

If the p-value is less than the usual $\alpha = 0.05$, then we have enough evidence to conclude that the independent variable differ significantly from zero. Hence it stays in the model.

2.3.4 Odds ratio

Re-writing equation (1) by taking the exponential of both sides of the equation, we have:

$$Odds = \frac{p}{1-p} = e^{\beta_0}e^{\beta_1 X_1}e^{\beta_2 x_2} \ldots e^{\beta_k X_k} \quad (10)$$

It is obvious from equation 5 above that when an independent variable X_i changes by 1 unit (all other variables kept constant), the odds changes by the factor e^{β_i}. This factor is termed the odds ratio (O.R) for the independent variable X_i. It gives the relative amount by which the odds of the outcome of interest increases (O.R > 1) or decreases (O.R < 1) when the value of the independent variable is changes by 1 unit.

3.0 Result and Discussion

In this section, details of model results and discussion are presented for each assertion. The results include model parameter values and statistics, the fitted Binary Logistic Regression Models, the -2 Log likelihood Null and Full model statistics, contingency table for Hosmer and Lemeshow test, and the tuberculosis treatment outcome classification table. The discussion is sectioned into the fitted TB good treatment outcome rate model, the goodness of fit of the fitted Bayesian Logistic Regression Model and the study implications to post harvest loss.

3.1 The fitted TB Binary Logistic Regression Model

Using the study data as captured by the TB treatment outcome (good or bad) as dependent variable, and the predictor variables (Sex (male or female), Treatment support (yes or no), HIV status (yes or no), Baseline sputum result (positive or negative), On Anti Retroviral Therapy (ART) (positive or negative) and On CPT drug (yes or no)), the Statistical Software for Social

Science (SPSS) was used to fit two (2) TB Binary Logistic Regression Model or what we refer to as the TB good treatment outcome prediction logistic models one each for both assertions. The Forward stepwise (Likelihood Ratio) method of regression was employed in fitting the models in three consecutive steps 1, 2 and 3 after the constant only model (step 0). The variable entered in each step of both models is found to be significant. Based on the first assertion that all the patients who successfully completed treatment were NOT cured, the final model (step 3) identified HIV status, Base line sputum and sex as significant predictor variables or model covariates ($p < 0.05$). See table 1 for details. In respect of the second assertion that all the patients who successfully completed treatment were cured, the final model (step 3) identified On ART, HIV status and treatment support as significant predictor variables or model covariates ($p < 0.05$). See table 2 for details. The Fitted TB good treatment outcome prediction model is given in the footnote of table 1 and 2 respectively for assertion 1 and 2.

The model regression parameters (β) their standard error (S.E), values of Wald's statistic and Exp(β) are also shown on tables 1 and 2 for each assertion respectively. For assertion 1, the covariate; HIV status, its corresponding value of Exp (β); 0.669 shows that if all other predictors are kept constant, the odds of good TB treatment outcome for HIV positive patients is 0.669 times its odds for HIV negative patients. Keeping all other predictors constant, the odds of good TB treatment outcome for TB patients whose sputum test positive at baseline is 54.938 times the odds of good treatment outcome for those whose sputum test negative at baseline. Still on this assertion, the covariate sex has the odd ratio 0.645. This means that while keeping all other predictors constant, odds of good TB treatment outcome for male patients, are 0.645 times the odds of good treatment outcome for female patients. See table 1 for details.

For assertion 2, the covariate On ART, with its Exp(β) value of 1.644, shows that the odds of good TB treatment outcome for patients on ART are 1.644 times the odds of good treatment outcome by those who are not on ART. The covariate HIV status has the odd ratio 0.592. This means that while keeping all other predictors constant, the odds of good TB treatment outcome for HIV positive patients, are 0.592 times the odds of good treatment outcome for HIV negative patients. Furthermore, the covariate treatment support has the odd ratio 0.759. This means that while keeping all other predictors constant, the odds of good TB treatment outcome for patients who had treatment support is 0.795 times the odds of good treatment outcome for patients who do not have. See table 2 for details.

Table 1: Binary Logistic Regression Model parameter values and statistics obtained based on assertion 11

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Base line sputum	4.044	.249	264.059	1	.000	57.045	35.026	92.905
	Constant	-3.602	.233	240.018	1	.000	.027		
Step 2[b]	Sex	-.397	.185	4.607	1	.032	.673	.468	.966
	Base line sputum	4.119	.253	264.573	1	.000	61.512	37.445	101.046
	Constant	-3.391	.249	184.828	1	.000	.034		
	Sex	-.438	.187	5.511	1	.019	.645	.448	.930

Step 3[c]	Base line sputum	4.006	.257	243.600	1	.000	54.938	33.219	90.858
	HIV Status	-.401	.176	5.184	1	.023	.669	.474	.946
	Constant	-3.114	.275	128.459	1	.000	.044		

Predictors in the final model: Baselinesp (Base line sputum), Sex, HIV Status.
TB Good treatment outcome rate $= \frac{1}{1+e^{-(3114-0.401\,HIV\ status\ +4.006\,Baselinesp-0.438Sex)}} * 100\%$, S.E = Standard error of model parameters, df = degree of freedom, Exp(β) = Odd ratio, C.I = confidence interval.

Table 2 : Binary Logistic Regression Model parameter values and statistics obtained based on assertion 2

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	HIVStatus	-.306	.123	6.158	1	.013	.737	.579	.938
	Constant	.956	.092	108.799	1	.000	2.600		
Step 2[b]	HIVStatus	-.537	.144	13.882	1	.000	.584	.440	.775
	ON_ART	.488	.163	8.986	1	.003	1.629	1.184	2.242
	Constant	.949	.092	107.202	1	.000	2.583		
Step 3[c]	Treatmentsupport	-.275	.140	3.862	1	.049	.759	.577	.999
	HIVStatus	-.525	.145	13.193	1	.000	.592	.446	.785
	ON_ART	.497	.163	9.282	1	.002	1.644	1.194	2.264
	Constant	1.010	.097	107.709	1	.000	2.745		

Predictors in the final model: ON_ART (On ART), HIV Status, Treatmentsupport
TB Good treatment outcome rate $= \frac{1}{1+e^{-(1.01+0.497ON\_ART\ -0.525HIVstatus-0.275Treatmentsupport)}} *$ 100%, S.E = Standard error of model parameters, df = degree of freedom, Exp(β) = Odd ratio, C.I = confidence interval.

Observe from the aforementioned result that, irrespective of the assertions – all successfully treated TB patients being cured or not, HIV status remains a predictor of TB treatment outcome with odd ratio less than one (OR < 1). This indicates a decrease in the odds of TB good treatment outcome for patients who test positive to HIV relative to that of those who test negative.

Observe also that if all the successfully treated TB patients are cured (assertion 2), then, sputum test result for TB at baseline and the sex of patient becomes immaterial in predicting TB treatment outcome. Rather the fact that patients are on treatment support and on ART become significant predictors of TB treatment outcome. Further observation reveal that, if all the successfully treated TB patients resisted treatment- not cured (assertion 1), then the predictors of TB treatment outcome will exclude the fact that patients are on treatment support and on ART but include their sputum test result at baseline and their sex. We state that the removal of outcome overlaps caused by the successful treatment outcome (comprising of patients who where cured and those who completed treatment) via the two assertions has made result interpretation a lot easier. This was not done by Hassan *et al*. (2006), making their result in our opinion, complex.

3.2 The goodness of fit of the fitted Bayesian Logistic Regression Model

The -2 Log likelihood Null and Full model summary on table 3 and 4 shows the change in the -2 Log likelihood and the significance at each step of the modeling process. This is respectively for assertion 1 and 2. For both assertions, step 3 (the last step), shows that the change is significant for all their respective model covariates; HIV status, Base line sputum and sex for assertion 1, and On ART, HIV status and treatment support for assertion 2 (p value < 0.05). This indicates a good fit for both models. See tables 3 and 4 below for details.

We also employ the contingency table for Hosmer and Lemeshow test, and the TB treatment outcome classification table for both assertions in establishing their respective model goodness of fit. The Hosmer and Lemeshow test statistic follows a chi-square distribution. The footnote of tables 5 and 6 shows for the respective assertions, that the model in step 3 is a good fit of the Bayesian Logistic Regression Model (p > 0.05). These contingency tables show close values of observed and expected frequencies for the good and bad Tb treatment outcome. These further buttress the goodness of the models. Respectively for the first and second assertions, the TB treatment outcome classification tables (tables 7 and 8) show 81.8% and 68.8% correct classification of the observed cases of TB treatment outcome by the model (see step 3 classification on tables 7 and 8) . This in addition establishes how good our model is.

Table 3: Change in -2 Log likelihood for each covariate at each step based on assertion 1

Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	p value of the Change
Step 1	Base line sputum	-735.453	580.623	1	.000
Step 2	Sex	-445.142	4.686	1	.030
	Base line sputum	-734.205	582.814	1	.000
	Sex	-443.030	5.620	1	.018
Step 3	Base line sputum	-691.254	502.069	1	.000
	HIVStatus	-442.798	5.157	1	.023

Table 4: Change in -2 Log likelihood for each covariate at each step based on assertion 2

Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	P value of the Change
Step 1	HIVStatus	-775.864	6.196	1	.013
Step 2	HIVStatus	-775.147	13.857	1	.000
	ON_ART	-772.767	9.096	1	.003
	Treatmentsupport	-768.219	3.814	1	.051
Step 3	HIVStatus	-772.895	13.165	1	.000
	ON_ART	-771.011	9.399	1	.002

Table 5: Contingency table for Hosmer and Lemeshow test based on assertion 1

		outcome = Bad outcome		outcome = Good outcome		Total
		Observed	Expected	Observed	Expected	
Step 1	1	697	697.000	19	19.000	716
	2	209	209.000	325	325.000	534
Step 2	1	406	405.808	9	9.192	415
	2	291	291.192	10	9.808	301
	3	159	159.192	222	221.808	381
	4	50	49.808	103	103.192	153
Step 3	1	256	258.048	7	4.952	263
	2	150	147.764	2	4.236	152
	3	211	212.676	8	6.324	219
	4	80	78.512	2	3.488	82
	5	59	55.495	55	58.505	114
	6	100	103.693	167	163.307	267
	7	50	49.812	103	103.188	153

Step 3 : χ^2 = 3.845, df = 5, p value = 0.572

Table 6: Contingency table for Hosmer and Lemeshow test based on assertion 2

		Treatment outcome = Bad		Treatment outcome = Good		Total
		Observed	Expected	Observed	Expected	
Step 1	1	225	225.000	431	431.000	656
	2	165	165.000	429	429.000	594
Step 2	1	128	128.728	195	194.272	323
	2	97	96.272	236	236.728	333
	3	165	165.000	429	429.000	594
Step 3	1	36	38.064	49	46.936	85
	2	92	90.702	146	147.298	238
	3	32	31.382	63	63.618	95
	4	41	39.877	82	83.123	123
	5	65	64.853	173	173.147	238
	6	124	125.123	347	345.877	471

Step 3 : χ^2 = 0.312, df = 4 , p value = 0.989

Table 7: Tuberculosis treatment outcome classification table based on assertion 1

Observed			Predicted		
			Treatment outcome		Percentage
			Bad outcome	Good outcome	Correct
Step 1	Treatment outcome	Bad outcome	697	209	76.9
		Good outcome	19	325	94.5
	Overall Percentage				81.8
Step 2	Treatment outcome	Bad outcome	697	209	76.9
		Good outcome	19	325	94.5
	Overall Percentage				81.8
Step 3	Treatment outcome	Bad outcome	697	209	76.9
		Good outcome	19	325	94.5
	Overall Percentage				81.8

a. The cut value is .500

Table 8: Tuberculosis treatment outcome classification table based on assertion 2

Observed			Predicted		
			Treatment outcome		Percentage Correct
			Bad outcome	Good outcome	
Step 1	Treatment outcome	Bad	0	390	.0
		Good	0	860	100.0
	Overall Percentage				68.8
Step 2	Treatment outcome	Bad	0	390	.0
		Good	0	860	100.0
	Overall Percentage				68.8
Step 3	Treatment outcome	Bad	0	390	.0
		Good	0	860	100.0
	Overall Percentage				68.8

a. The cut value is .500

4.0 Conclusion and Recommendations

4.1 Conclusion

The following conclusions were drawn from the study:

(i) Logistic modeling of tuberculosis treatment outcomes among farmers using response variable outcomes from non- disjoint events as being successfully done in this study without complications in result interpretation

(ii) Whether all successfully treated TB patients are cured or not, HIV status remains a predictor of TB treatment outcome with decrease in the odds of TB good treatment outcome for patients who test positive to HIV relative to that of those who test negative.

(iii) If all the successfully treated TB patients are cured, sputum test result for TB at baseline and the sex of patient becomes immaterial in predicting TB treatment outcome. Rather, the fact that patients are on treatment support and on ART become significant predictors of TB treatment outcome with odd ratios of 1.644 and 0.759 respectively.

(iv) If all the successfully treated TB patients resisted treatment (not cured), the predictors of TB treatment outcome excludes the fact that patients are on treatment support and on ART but include their sputum test result at baseline and their sex with odd ratios of 54.938 and 0.645 respectively.

4.2 Recommendations

The following recommendation was made in the study:

In a further research, the Binary Logistic Model should be fitted for varying proportions of patients who completed treatment and were cured or not cured in order to identify significant predictors of good treatment outcome.

References

Coasta-Veiga, A., Teodoro, B. and Nunes, C. (2018). Unsuccessful treatment in pulmonary tuberculosis: factors and a consequent predictive model. *European Journal of Public health* 28 (2): 352-358.

Getahun, B., Ameni,G., Medhim, G., and Biadgilign, S. (2013). Treatment outcome of tuberculosis patients under directly observed treatment in Addis ABABA, Ethiopia. The *Brazilian journal of infectious diseases* 17 (5) : 521-528.

Hassan, A., Olukolade, R., Ogbuji, Q., Onyemocho, A., Okwuonye, L., Igbabul, S., Okechukwu, J., Kusimo, O., Osho, A., Osinowo, K., and Ladipo, O. (2016). Evaluation of tuberculosis treatment outcome of TB/HIV Co- Infection: A four – year retrospective cohort study in HIV – prevalent setting of north central Nigeria. *Journal of tuberculosis research* 4 (3): 122-133.

Mohammed, M. E. and Alnory, A. (2020). Logistic regression analysis to determine cardiovascular diseasesrisk factors. A hospital – based case control study, 2019. *International journal of medical science and clinical invention* 7 (8): 4951-4959.

Mohamed, R . A. and Mohamed, G. I. (2018). On the estimation methods for binary logistic regeression model with missing values. *International Journal of Mathematical and Computational Science* 4 (3): 79-85.

Peter, S. K. T. (2020). Determinants of unsuccessful treatment outcomes and mortality tuberculosis patients in Malasia: A registry based cohort study. *PLoS One* 15 (4) : e0231986.

Peduzzi P., Concato J., Kemper E., Holford, T. R., and Feinstein A. R. (1996). A Simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemology* 49:1373-1379.

Sharareh,R., Niakan, K., Mahshid, N., and Xiano-Jun, Z. (2010). A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion. *IJAM* 40 (2): 2-8.

Teshome, K. A., and Anagaw, Y.K. (2017). Outcome of tuberculosis treatment and its predictors among HIV infected patients in southwest Ethiopia. *International journal of General Medcine* 10: 161-169.