International Journal of Scientific Engineering and Technology ISSN:2277-1581
Volume No.6 Issue No.2, pp: 99-104 01 Feb. 2017

Binary Logistic Regression Model Application: Identification of Factors Associated to MTCT of HIV in Rwanda

Cassien Habyarimana[1,]*; Alexandre Lyambabaje[2]
[1]Integrated Polytechnic Regional Centre, Rwanda; [2]Inter-University Council for East Africa (IUCEA)
*Corresponding author. Email address: habyarimanacassien@gmail.com

Abstract: In this paper, we estimated the HIV infection prevalence and identified the factors associated to HIV free survivals. We have found that the mother's current HIV status and the mother's age are predictor variables which have a significant impact on HIV infection; they are factors significantly associated to MTCT of HIV in Rwanda.
Keywords: Binary Logistic Regression, HIV Infection, Prevalence, Odd Ratio.

1. Introduction

The HIV epidemic continues to be one of the major global health challenges facing the world today. According to (UNAIDS, 2009), an estimated 33.4 million people are living with HIV/AIDS worldwide, with 2.1 million children infected. In sub-Saharan Africa alone, about 22.4 million people (67% of HIV infections worldwide) are infected, of which 60% are women, thereby jeopardizing the health of future generations. According to the same report, 71% of the new HIV infections and 72% of the estimated number of deaths due to AIDS in the world occurred in Sub Saharan Africa. Additionally, the number of HIV-infected children younger than 15 years increased from 1.6 million in 2001 to 2.0 million in 2007, with 90% of these children living in Sub-Saharan Africa (UNAIDS, 2007). Every day, more than 8000 people die of AIDS globally, mainly because of a lack of access to prevention and treatment services. In highly affected countries, entire families and communities have been decimated by HIV and AIDS which in turn threaten the socioeconomic fabric of these countries. Progress has been made in the introduction and scaling up of prevention of mother to child HIV transmission programs (PMTCT), which allows HIV positive women to have access to services that can improve their health and prevent HIV transmission to their infants. The United Nations strategic approach to PMTCT is a four prongs strategy which includes: primary prevention of HIV infection; prevention of unintended pregnancies among HIV-infected women; prevention of HIV transmission from HIV-infected women to their infants; and the provision of care and support to HIV-infected women and their infants and families. While the rate of vertical transmission from mother to child has been reduced to below 2% in industrialized countries through adequate prophylaxis, elective c-section and replacement feeding, paediatric AIDS remains an uncontrolled epidemic in developing countries (L Creek, et al., 2008), even though the number of new infections in infants has decreased from 460,000 to 420,000. HIV transmission to infant can occur in womb (8-10%), during peri-partum period (10-15 %), and post-partum through breastfeeding (15-20%). Without any prevention, vertical transmission rates vary from 15% to 25% in industrialized countries; and from 25% to 45% in developing countries (UNAIDS, 1999).

While progress has been made in terms of reducing HIV infections in children, a valid measure of PMTCT program effectiveness is lacking. In 2005, only 2% of exposed infants globally received an HIV test in their first 18 months of life, which complicates the evaluation of PMTCT effectiveness (Luo C, et al., 2007). Several programs to prevent the transmission of HIV from mother to child are implemented across Rwanda. However, little is known about their real impact and limiting factors. In Rwanda, different Partners including Global Fund, UNICEF, and the School of Public Health are involved in research activities which are focusing on the identification of socio-demographic and PMTCT factors associated to HIV transmission from mother to child. The logistic regression model allows for the assessment of the most significant ones and fits the relationship between the independent variables and the probability that the dependent variable occurs (Agresti, A., 1996).

2. Logistic Regression Model

2.1. Model

The logistic model formula computes the probability of the selected response as a function of the values of the explanatory variables.
The model has the following form,

$$\pi(X) = \Pr(Y = 1 \mid X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}} \qquad (1)$$

Our model will be predicting the logit, which is the natural log of the odds of having made one or the decision. That is,

$$\text{logit}(\pi(X)) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \qquad (2)$$

where $\beta_0, \beta_1, \beta_2, ..., \beta_k$ are known as the regression coefficients, which have to be estimated from the data; and Y be a binary response variable denoting the outcome of some experiment (Hosmer, David W. and Stanley Lemeshow, 2000).

2.2. Odds and Odds Ratios

2.2.1. Odds

The "odds or chance" of an event is defined as the probability of the outcome event occurring divided by the probability of the event not occurring.

Note that, for our model

$$\ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$; the quantity $\frac{\pi(X)}{1-\pi(X)}$ is the *Odds* of the Outcome of interest for an individual with covariates X, $\ln\left(\frac{\pi(X)}{1-\pi(X)}\right)$ is called the *Log Odds* or *logit*.

The logistic regression estimates the probability of a certain event occurring and thus forms a predictor variable $\ln\left(\frac{\pi(X)}{1-\pi(X)}\right)$ which is a linear combination of the explanatory variables. The values of this predictor variable are then transformed into probabilities by a logistic function.

2.3.2 Odds ratios

Logistic regression also produces Odds Ratios (OR) associated with each predictor value. The Odds ratios associated with X_j is given by,

$$OR_{X_j} = \frac{\text{Odds}(\cdots, X_j = x_j + 1, \cdots)}{\text{Odds}(\cdots, X_j = x_j, \cdots)} = e^{\beta_j}$$ independent of X_j.

If $X_j \in \{0, 1\}$ is dichotomous, then odds for group with $X_j = 1$ are e^{β_j} higher, other parameters being equal.

In general, the "*odds ratio*" is one set of odds divided by another. The odds ratio for a predictor is defined as the relative amount by which the odds of the outcome increase (OR>1) or decrease (OR<1) when the value of the predictor variable is increased by 1 unit (Hilbe, Joseph M., 2009).

3. ESTIMATION

3.1. Likelihood Function, Maximum Likelihood Estimation

Consider a random sample $Y_1, Y_2, \ldots, Y_n$ from the Bernoulli distribution $\Pr[Y_j = y] = p_0^y (1-p_0)^{1-y}$, for $y = 0,1$ where p_0 is an unknown probability. *The question is how to estimate* p_0?

Next, let $y_1, y_2, \ldots, y_n$ be a given sequence of zeros and ones. Thus, each y is either 0 or 1. Hence,

$$f_n(y_1, y_2, \ldots, y_n \mid p_0) = p_0^{\sum_{j=1}^n y_j} . (1-p_0)^{n - \sum_{j=1}^n y_j} .$$

Replacing the given non-random sequence $y_1, y_2, \ldots, y_n$ by the random sample $Y_1, Y_2, \ldots, Y_n$ and the unknown probability p_0 by a variable p in the interval (0,1) yields the Likelihood function

$$L_n(p) = f_n(Y_1, Y_2, \ldots, Y_n \mid p) = p^{\sum_{j=1}^n Y_j} . (1-p)^{n - \sum_{j=1}^n Y_j}$$

For the case $p = p_0$ the likelihood function can be interpreted as the joint probability of a particular sample $Y_1, Y_2, \ldots, Y_n$. (Miller, R. et. Al., 1980).

Suppose in a population from which we are sampling, each individual has the same probability, p, that an event occurs. For each individual in our sample of size *n*, $Y_j = 1$ indicates that an event occurs for the i[th] subject, otherwise, $Y_j = 0$.

The likelihood function is given by,

$$L_n(p) = p^{\sum_{j=1}^n Y_j} . (1-p)^{n - \sum_{j=1}^n Y_j}$$

The log of this equation is called the *log likelihood function*, and is defined as,

$$l_n(p) = \ln(L_n(p)) = \sum_{j=1}^n Y_j \ln(p) + (n - \sum_{j=1}^n Y_j).\ln(1-p).$$

For estimation, we will work with the log-likelihood function.

The maximum likelihood estimate (MLE) of p is the value that maximizes l_n (equivalent to maximizing L_n). Let the sample mean be $\bar{Y} = \frac{1}{n}\sum_{j=1}^n Y_j$. We get

$$\ln(L_n(p)) = n.\left[\bar{Y}\ln(p) + (1 - \bar{Y})\ln(1-p)\right]$$

Therefore, the ML estimator $\hat{p}$ can be obtained from the first-order condition for a maximum of $\ln(L_n)$ in $p = \hat{p}$. That is,

$$0 = \frac{d\ln(L_n(\hat{p}))}{d\hat{p}} = n\left[\frac{\bar{Y} - \hat{p}}{\hat{p}(1-\hat{p})}\right]$$, and from this we get $\hat{p} = \bar{Y}$. The maximum likelihood estimator $\hat{p}$ of p is the sample mean, $\hat{p} = \frac{1}{n}\sum_{j=1}^n Y_j$ (Hosmer, D. W., Jr. and Lemeshow, S., 1989; Miller, R. et. Al., 1980).

3.2. Maximum Likelihood Estimation for the Coefficients in Logistic Regression model

The equation (1) can be written as, $\pi = \frac{\exp(X\beta)}{1 + \exp(X\beta)}$

where $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ and $X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$

is called the regression (or predictor) matrix. For the i[th] observation, we write

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik})}.$$

The logistic regression problem is then to obtain an estimate of the vector $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$.

The method used to find the parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$ is the method of maximum likelihood. Specifically, $\hat{\beta}$ is the value that maximizes the likelihood function,

$$L_n(\pi_i) = \prod_{i=1}^n \pi_i^{y_i}\left[1 - \pi_i\right]^{1-y_i} = \prod_{i=1}^n \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}(1-\pi_i).$$

The log of this equation is called the log likelihood, and is defined as

$$l(\beta) = \ln(L(\beta)) = \sum_{i=1}^{n}\left[y_i \ln(\pi_i) + (1-y_i)\ln(1-\pi_i)\right].$$

Substituting the value of π_i in this equation, we get (with $x_{i0}=1$),

$$\ln(L_n(\beta)) = \sum_{i=1}^{n}\left[y_i \cdot \sum_{j=0}^{k} x_{ij}\beta_j - \ln\left(1+\exp\left(\sum_{j=0}^{k} x_{ij}\beta_j\right)\right)\right] \quad (3)$$

To find the critical points of the log likelihood function, set the first derivative with respect to each β equal to zero.

Thus, differentiating equation (3) with respect to each β_j,

$$\frac{\partial \ln(L_n(\beta))}{\partial \beta_j} = \sum_{i=1}^{n}\left[y_i x_{ij} - \pi_i x_{ij}\right] \quad (4)$$

The maximum likelihood estimates for β can be found by setting each of the k+1 equations in equation (4) equal to zero and solving for each β_j.

The general form of the matrix of second partial derivative is:

$$\frac{\partial^2 \ln(L_n(\beta))}{\partial \beta_j \partial \beta_{j'}} = \frac{\partial}{\partial \beta_{j'}}\left(\sum_{i=1}^{n}\left[y_i x_{ij} - \pi_i x_{ij}\right]\right)$$

$$= -\sum_{i=1}^{n}\left[x_{ij}\pi_i(1-\pi_i)x_{ij'}\right] \quad (5)$$

The values of all β_j satisfying the conditions of equation (4) and (5) are called Maximum Likelihood Estimators of β_j noted by $\hat{\beta}_j$. We deduce a **MLE** of logistic regression function (Albert A. and Anderon, J.A., 1984),

$$\hat{\pi}_i = \frac{\exp\left(\sum_{j=0}^{k} x_{ij}\hat{\beta}_j\right)}{1+\exp\left(\sum_{j=0}^{k} x_{ij}\hat{\beta}_j\right)}.$$

3.3. Parameter Variances and Covariances

Estimates for the variances and covariances of the estimated parameters $\hat{\beta}_j$ are computed as follows.

Let $\hat{I}(\hat{\beta}) = X'VX$, where X is the $n \times n$ regression matrix, and V is an $n \times n$ diagonal matrix with ith diagonal term $\hat{\pi}_i(1-\hat{\pi}_i)$. Denote, $\hat{\Sigma}(\hat{\beta}) = \hat{I}^{-1}(\hat{\beta})$. The estimate of the variance of $\hat{\beta}_j = \hat{\sigma}^2(\hat{\beta}_j)$ is then the jth diagonal term of the matrix $\hat{\Sigma}(\hat{\beta})$, and the off-diagonal terms are the covariance estimates $\sigma(\hat{\beta}_j, \hat{\beta}_k)$ for $\hat{\beta}_j$ and $\hat{\beta}_k$ (Freeman, D. H., Jr., 1987).

3.4. Significance of the model

In this section, the three measures of goodness of fit; the G statistic, Pearson statistic, and Hosmer-Lemeshow statistic are discussed.

3.4.1. G Statistic

The G statistic measures a difference in deviance between two models. For logistic regression, the deviance of a model is defined as follows

$$D = -2\sum_{i=1}^{n}\left[Y_i \ln\left(\frac{\hat{\pi}_i}{Y_i}\right) + \left(1-Y_i\right)\ln\left(\frac{1-\hat{\pi}_i}{1-Y_i}\right)\right].$$

The larger the difference, the greater the evidence that the model is significant.

3.4.2. Pearson Statistic

For the Pearson statistic, the predictor matrix rows are placed into J groups such that identical rows are placed in the same group. Then the Pearson statistic is given by

$$p = \sum_{j=1}^{J}\frac{(O_j - m_j\hat{\pi}_j)^2}{m_j\hat{\pi}_j(1-\hat{\pi}_j)}$$

where O_j is the number of positive observations for group j, $\hat{\pi}_j$ is the model's predicted value, and m_j is the number of identical rows.

3.4.3. Hosmer-Lemeshow Statistic

The Hosmer-Lemeshow statistic takes an alternative approach to grouping: it groups the predictions of a logistic regression model rather than the model's predictor variable data, which is the Pearson statistic's approach. Model predictions are split into G bins that are filled as evenly as possible. Then the statistic is computed as follows

$$HL = \sum_{j=1}^{G}\frac{(O_j - n_j\bar{\pi}_j)^2}{n_j\bar{\pi}_j(1-\bar{\pi}_j)}$$

where O_j is the number of positive observations in group j, $\bar{\pi}_j$ is the model's average predicted value in group j, and n_j is the size of the group (Menard, Scott., 1995).

3.5. Confidence Intervals and Hypothesis Testing

3.5.1. Confidence intervals for the coefficients of the model

In the model

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k,$$

a (1-α)x100% confidence interval for β_j, j = 1,2,...,k, can easily be calculated as

$$\left[\hat{\beta}_j - Z_{1-\frac{\alpha}{2}}.\hat{\sigma}(\hat{\beta}_j) \quad , \quad \hat{\beta}_j + Z_{1-\frac{\alpha}{2}}.\hat{\sigma}(\hat{\beta}_j)\right]$$

where $\hat{\beta}_j$ is the Maximum Likelihood Estimator of β_j and $\hat{\sigma}(\hat{\beta}_j) = \hat{\pi}_j(1-\hat{\pi}_j)$.

3.5.2. Confidence intervals for the odds ratio

In the model

ISSN:2277-1581
01 Feb. 2017

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k ,$$

a (1-α)x100% confidence interval for the odds ratio over a one units change in X_j is

$$\left[\exp\left(\hat{\beta}_j - Z_{1-\frac{\alpha}{2}}.\hat{\sigma}(\hat{\beta}_j)\right), \exp\left(\hat{\beta}_j + Z_{1-\frac{\alpha}{2}}.\hat{\sigma}(\hat{\beta}_j)\right)\right]$$

where $\hat{\beta}_j$ is the Maximum Likelihood Estimator of β_j and $\hat{\sigma}(\hat{\beta}_j) = \hat{\pi}_j(1 - \hat{\pi}_j)$.

3.5.3. Hypothesis testing of significance of all predictors

Let us consider the model of equation (2). We want to test the null hypothesis that all coefficients $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are simultaneously zero, i.e., we have to test the following hypothesis:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1 : \text{At least one of the coefficien ts } \beta_1, \beta_2, \cdots, \beta_k \text{ is not zero.} \end{cases}$$

This is like the overall F-test in linear regression. In other words, this is to testing the null hypothesis that an intercept-only model is correct,

$$\ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0$$

versus the alternative that the current model is correct,

$$\ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

This test has k degrees of freedom, that is the number of β-parameters (except the intercept β_0). Large chi-square statistics lead to small p-values and provide evidence against the intercept-only model in favor of the current model (G. Rodriguez, Revised September 2007).

3.6. Interpretation of Logistic Regression Coefficients and Odds Ratios

Logistic regression coefficients are interpreted as follows,

- $\exp(\beta_0)$ = the odds that the characteristic is present in an observation *i* when $X_i = 0$
- $\exp(\beta_j)$ = for every unit increase in X_{ij}, the odds that the characteristic is present is multiplied by $\exp(\beta_j)$. This is an estimated odds ratio.

$$\frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_j (X_{ji} + 1) + \beta_k X_{k,i})}{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{k,i})} = \exp(\beta_j)$$

In general, the logistic regression model stipulates that the effect of a covariate on the chance of "success" is linear on the log-odds scale, or multiplicative on the odds scale.

- If $\beta_j > 0$, then $\exp(\beta_j) > 1$, and the odds increase.
- If $\beta_j < 0$, then $\exp(\beta_j) < 1$, and the odds decrease (Hilbe, Joseph M., 2009).

4. Data Analysis, Results and Discussion

4.1. Introduction

The data used for this paper are from a national household survey targeting children born to HIV+ mothers. Data have been collected from East, North, West, South and Kigali city. A sample of 3020 women was selected. The independent variables analyzed are: "(1)mother's age, (2)mother's marital status, (3)mother belongs to an association, (4)has electricity at home, (5)has a car, (6)mother had ever sex with infected persons, (7) mother's Current HIV status", and the dependent variable were "child's HIV status at time of survey". The significance of the above independent variables was tested using SPSS.

4.2. Descriptive Statistics for predictor variables

Variable	Frequency	Percentage (%)
Marital Status		
Single	223	7.4
Cohabit	719	23.8
Married	1636	54.2
Divorced	209	6.9
Widowed	214	7.1
Not Applicable (NA)	19	0.6
Total	**3001**	**100.0**
Belong to an association		
Yes	1436	47.5
No	1491	49.4
NA	93	3.1
Total	**3020**	**100.0**
Access to electricity		
Yes	117	3.9
No	2885	95.5
NA	18	0.6
Total	**3020**	**100.0**
Access to a car		
Yes	11	0.4
No	2983	98.8
NA	26	0.9
Total	**3020**	**100.0**
Sex with infected people		
Yes	2557	84.7
No	452	15.0
NA	11	0.4
Total	**3020**	**100.0**
Mother's HIV status		
Don't know	212	7.0
HIV-	1326	43.9
HIV+	20	0.7
NA	1462	48.4
Total	**3020**	**100.0**

Table 1: Frequency table for qualitative predictors

From **Table 1**, 40.7% of the total number of mothers participated are aged 25 to 34, while 54.2% of them are married. The number of women who belong to the associations is significantly equally likely that who do not belong to the associations. 95.5% of total sample has no access to electricity and 98.8% of them have no own cars. This category is more considered than the one of who have no own cars.

A significant number of women (84.7%) heard sex with infected person. The mother's HIV status shows that 43.9% are infected while 48.4% didn't respond on their status.

	N	Minimum	Maximum	Mean	Std. Deviation
Age	3007	11	50	31.13	6.296
Valid N (listwise)	3007				

Table 2: Descriptive statistics for mother's age

The mother's age is 31 in average but with minimum of 11 and maximum of 50 years old (see **Table 2**).

4.3. Binary Logistic Regression results

4.3.1. Hypothesis Testing for Data Sample, and dependent variable encoding

Assume that X_1= "mother's age" , X_2= "Marital status of the participants", X_3= "Participants involvement in Association", X_4= "Access to electricity", X_5 = " Access to car" , X_6 = " Mother had sex with infected people" and X_7 = "Mother's HIV status", and use the following coding for dependent variable

Original Value	Internal Value
Negative	0
Positive	1

I have coded the child's HIV status with 0="Negative" and 1="Positive"

Table 3: Dependent variable encoding

Our regression model will be predicting the logit (natural log of the odds) of being tested negative or positive. That is,

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\beta_4 X_4+\beta_5 X_5+\beta_6 X_6+\beta_7 X_7,$$

where π is the predicted probability of the event which is coded with 1 (tested Positive), $1-\pi$ is the predicted probability of being tested negative, and $X_1,...,X_7$ are our predictor variables.

The hypothesis test is

H_0 : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$,

H_1 : At least one of the $\beta_j \neq 0$, for j = 1, 2,…. ,7.

4.3.2. Interpreting Parameter Estimates

Look at the statistical output. We see that there are 1407 cases used in the analysis.

Unweighted Cases		N	Percentage
Selected Cases	Included in Analysis	1407	46.6
	Missing Cases	1613	53.4
Unselected Cases		0	0.0
Total		**3020**	**100.0**

Table 4: Case Processing Summary

The **Block 0** output is for a model that includes only the intercept (constant). Given the base rates of the two test results (1403/1407=99.7% tested negative, 0.3% tested positive).

Observed			Predicted		
			HIV status		Percentage Correct
			HIV -	HIV +	
Step 0	HIV status	HIV -	1403	0	100.0
		HIV +	4	0	0.0
Overall Percentage					99.7

Table 5: Classification table (Block 0)

Looking at Table 4, we found that the prevalence of HIV is 0.2% (4/1407) in Rwanda.

4.3.3. Model Validation

Variable	B	S.E.	Wald	df	Sig.	Exp(B)
Mother's age	-1.850	1.326	1.948	1	0.163	0.157
Mother's HIV status			19.068	2	0.000	
HIV -	-5.866	1.378	18.109	1	0.000	0.003
HIV +	-3.908	1.369	8.150	1	0.004	0.020
Constant	-0.880	1.013	0.721	1	0.396	0.423

Table 6: Variables in the equation

The **Variables in the Equation** output for its last step (**Table 6**) shows that:

- The mother's Current HIV status has a highly significant effect on the child's HIV status at time of survey because the associated p-value is sig = 0.000 < α = 0.10. In this case the HIV- status referenced by HIV+ status is very high significantly rather than the Don't know referenced by HIV+ status.
- The mother's age also has a small effect on the child's HIV status at time of survey because the associated p-value is sig = 0.163 ≈ 0.10.
- With significantly associated factors to the child's HIV status, the predictive model is
-

$$\ln(\text{ODDS}) = \ln\left(\frac{\pi}{1-\pi}\right) = -0.860 - 1.850\ X_1 - 5.886\ X_7.$$

Concluding remarks

In this paper, the factors associated to mother to child transmission of HIV, are identify, and the most important and significant variable is the mother's HIV status during pregnancy. The mother's age has a small quasi-significant effect on child's HIV status and other independent variables have no significant effect on the child's HIV status**.**

 ISSN:2277-1581

References

i. *Agresti, A. (1996), An Introduction to Categorical Data Analysis, New York: Wiley.*

ii. *Albert A. and Anderon, J.A. (1984), "On the existence of maximum likelihood estimates in logistic regression models". Biometrika, 71.*

iii. *Freeman, D. H., Jr. (1987), Applied Categorical Data Analysis, New York: Marcel Dekker, Inc.*

iv. *G. Rodriguez, Revised September 2007. Logit Models for Binary Data.*

v. *Hilbe, Joseph M. (2009). Logistic Regression Models. Chapman & Hall/ CRC Press*

vi. *Hosmer, D. W., Jr. and Lemeshow, S. (1989), Applied Logistic Regression, New York: John Wiley & Sons, Inc.*

vii. *Hosmer, David W. and Stanley Lemeshow (2000). Applied logistic Regression, 2nd Ed. New York; Chichester, Wiley.*

viii. *L Creek, et al., (2008). Monitoring Effectiveness of Programmes to Prevent Mother-to-child HIV Transmission in Lower-Income Countries. WHO Bulletin, Volume 86, Number 1, January 2008, 1-80.*

ix. *Luo C, et al., (2007). Global Progress in PMTCT and Paediatric HIV Care and Treatment in Low- and Middle-income Countries in 2004-2005. Reprod Health Matters. 2007 Nov; 15(30):179-89.*

x. *Menard, Scott. (1995). Applied Logistic Regression Analysis. Sage Publication. Series. Quantitative Applications in the Social Sciences, No. 106.*

xi. *Miller, R. et. Al., (1980), Biostatistics Casebook, john Wiley & Sons.*

xii. *UNAIDS (1999). Report on the global AIDS epidemic. Geneva, UNAIDS.*

xiii. *UNAIDS (2007).Practical guidelines for intensifying HIV prevention. Geneva, UNAIDS.*

xiv. *UNAIDS (2009). Report on the global AIDS epidemic. Geneva, UNAIDS.*