Bivariate Response Logistic Regression for Categorical Data

***Adah, V., and Nwaosu, S. C.,**
Department of Mathematics/Statistics/Computer Science,
University of Agriculture, Makurdi, Benue State.

Nja, M. E.,
Department of Statistics, University of Calabar, Cross River State.
*Corresponding author's email:adahvictoria14@yahoo.com

Abstract

The bivariate logistic regression model can be used to obtain the probability of joint events as well as individual events where there are two response variables and several explanatory variables. The existing bivariate logistic model approach appears intractable. This paper provides a modeling procedure that addresses this problem. This approach compares favourably with the existing procedure. The new approach is used to model the probability of malaria and typhoid infections, using age, sex and location of the patients as associated factors. The marginal probabilities showed a decrease in malaria infection with age. Sex and location showed a significant impact on the probability of malaria infection. Typhoid fever infection on the other hand indicates an increase with age. Sex has no significant impact on the probability of typhoid infection. The joint model shows that all variables are statistically significant with odds value greater than 1 indicating higher likelihood of joint infection and odds value that are less than one indicating lower likelihood of joint infections, $\chi 2$:12.02828 (0.00729)

Keywords: Bivariatebinary, odds ratio, response probability, marginal model, joint model

Introduction

The Bivariate logistic regression characterizes the dependence of each response variable on explanatory variable and describes the association between the (outcome) response variables (Liang and Qaqish, 1994). Dependence in outcome variables may arise in various fields including epidemiology (Islam, 2013). Islam *et al.* (2013), applied the bivariate logistic regression model to the Health and Retirement Study (HRS) data. Multivariate logistic regression analysis of complex survey data with application to behavioral risk factor surveillance system (BRFSS) data was examine by Lu and Yang (2012). Li and Wong (2011) used the logistic regression to examine the presence of multiple types of decompression sickness (DCS) in animal. In this study the *Bivariate* logistic regression is used to model two binary response variables (malaria and typhoid).

Malaria is an infectious disease caused by a parasite, Plasmodium, which infects human red blood cells (RBC's) (Francis, 2010). It is transmitted by female *Anopheles* mosquitoes and characterized by cycles of chills, fever, pain and sweating (Pradhan, 2011). It makes a person weak as frequent attacks of malaria destroy the RBC's in the blood. As a result, it makes the person anaemic and prone to other infections. Malaria is one of the leading causes of disease and death in children and adults in countries where it is endemic (WHO, 2015a). Although the disease has been eradicated in most temperate zones, it continues to be endemic throughout most of the tropics and subtropics (Hagenlocher, 2015).

Typhoid fever (enteric fever) on the other hand is a systemic prolonged febrile illness caused by certain Salmonella serotypes. Poor disposal of human excreta, poorly equipped latrine with water facility, poor hand washing habit, and untreated water usage are the main cause of transmission of typhoid fever in developing countries (Birhanie *et al*, 2014). Like malaria, typhoid fever is belief to be endemic and prevalent in Nigeria (Igbeneghu, 2009).

Although malaria and typhoid are caused by very different organisms - one a protozoan, the other a gram negative bacilli, and transmitted via different mechanisms, both diseases share rather similar symptoms (Birhanie *et al*, 2014).

An association between malaria and typhoid fever was first described in the medical literature in the middle of the 19th century, and was named typho-malarial fever by the United States Army (Pradhan, 2011). High prevalence of malaria is an established fact; it is only within the last few decades that an unusually high number of illnesses have been diagnosed as malaria coexisting with typhoid fever (Pradhan, 2011).

However, the precise incidence of the concurrent malaria and typhoid fever in most geographical areas is largely uncertain. Typhoid and malaria share social circumstances which are imperative to their transmission, hence, individuals in areas endemic for both diseases are at substantial risk of contracting both diseases, either concurrently or an acute infection superimposed on a chronic one (Keong and Sulaiman, 2006). Both diseases have been associated with poverty and underdevelopment with significant morbidity and mortality (Uneke, 2008). Igiri *et al* (2018) studied malaria-typhoid co-infection among patients in Ahmadu Bello University Teaching Hospital Zaria, Kaduna State, Nigeria. The study shows substantial association exists in the malaria-typhoid fever co-infection between the age groups. There was a statistical relationship between malaria and typhoid fever at (r = 0.967, $p > 0.05$). Jones *et al* (2015),Yury*et al* (2015) used the determinist models to study the malaria and typhoid fever co-infection. Logistic regression has been widely used in epidemiological studies to identify relevant predictors of Malaria and Typhoid based on marginal approaches for instance Agomo and Oyibo (2013), Okafor and Oko-Ose (2012), Kazembe *et al*. (2006), Hagenlocher and Castro (2015).The association of Malaria and Typhoid fever and its risk factors is of great importance. The study is immensely significant as it was used to model the probability of the joint infection of Malaria and Typhoid fever, which is the gap in literature this study seeks to fill.

This study makes use of secondary data collected from the General hospital Otukpo, Benue State for the year 2014.The predictor variables used are the age, sex and location of the patients.

Methods

Model Specification

The bivariate logistic regression is used to provide parameter estimation for

The marginal probability of malaria infection.

The marginal probability of typhoid fever infection

The probabilities for joint association of malaria and typhoid fever.

For the i-th (i= 1, 2…n) individual response the marginal probabilities for malaria and typhoid fever infections are given by the regression model:

Model 1:

$$logit(\mu_{ij}) = B_0 + B_1X_1 + B_2X_2 + B_3X_3 \quad j = 1,2 \quad (1)$$

where:

B_0 = is the Y intercept

B_s; s = 1,2,3 are the slope parameters.

X_1 = The age of the patient$(0.5 < 4, 4 < 9, 9 < 15, \geq 15)$

X_2 = Sex(Male/Female)

X_3 = Location(Urban/Rural)

$$\mu_{ij} = \frac{e^{B_0 + B_1X_1 + B_2X_2 + B_3X_3}}{1+e^{B_0 + B_1X_1 + B_2X_2 + B_3X_3}}, \text{ j= 1,2}$$

μ_{i1} is the response probability of malaria infection

μ_{i2} is the response probability of typhoid fever infection.

Bivariate Logistic Regression

Bivariate logistic regression (BLR) differs from ordinary logistic regression in the sense that response variables are two (Y_1 and Y_2). If multiple binary outcomes are assessed on the same individual, they share the same characteristics to that individual and are likely to exhibit correlations within the subject (Lu and Yang, 2012. In this regard, bivariate logistic regression is a useful procedure with advantages that include:

Individual modeling of the marginal probability distribution of the bivariate binary responses.

Modeling the pairwise association between the two binary responses in relation to several covariates (Yee, 2008).

The parameter estimation is by Maximum likelihood method, McCullagh and Nelder (1989) and Le Cessie and Houwellingen (1994). The pairwise association between the two binary responses is estimated in terms of odds ratio and hence the joint probabilities.

The drawback of such estimation procedure is that the joint model for each of the possible probabilities cannot be estimated easily (That is, the probability that a person has malaria and/or typhoid). Thus it seems better to have joint models that use an estimating procedure where each of these probabilities can be obtained as:

For the *i*-th individual response (*i*=1, 2…n) the variables (Y_{1i} and Y_{2i}) are the indicator variable for Malaria and Typhoid fever respectively. That is,

$$Y_{1i} = \begin{cases} 1, & \text{if } i \text{ has Malaria} \\ 0, & \text{otherwise} \end{cases}$$

$$Y_{2i} = \begin{cases} 1, & \text{if } i \text{ has Typhoid} \\ 0, & \text{otherwise} \end{cases}$$

A regression model expressing the association between these two responses is given by:

$$\text{Model}_{ij}\text{: logit } (\mu_{ijk}) = \boldsymbol{\eta}_{ijk}(\text{x})$$

$$= \beta_{ij0} + \beta_{ij1}X_1 + \beta_{ij2}X_2 + \beta_{ij3}X_3; i,j = 0,1 \quad (2)$$

$$\boldsymbol{\beta_{11}} = [\beta_{110}, \beta_{111}, \beta_{112}, \beta_{113}],$$

$$\boldsymbol{\beta_{10}} = [\beta_{100}, \beta_{101}, \beta_{102}, \beta_{103}],$$

$$\boldsymbol{\beta_{01}} = [\beta_{010}, \beta_{011}, \beta_{012}, \beta_{013}],$$

$$\boldsymbol{\beta_{00}} = [\beta_{000}, \beta_{001}, \beta_{002}, \beta_{003}],$$

$$\mathbf{X'_i} = [1, X_{1i}, X_{2i}, X_{3i}]$$

The response from an individual patient is therefore occurring with fourpossible probabilities:

$$\mu_{ij} = \frac{e^{B_{ij}X'_i}}{1 + e^{B_{ij}X'_i}} \quad (3)$$

μ_{11}: The response probability of a patient having Malaria and Typhoid fever.

μ_{01}: The response probability of a patient not having Malaria but Typhoid fever.

μ_{10}: The response probability of a patient having Malaria but not Typhoid fever.

μ_{00}: The response probability of a patient having neither Malaria nor Typhoid fever.

The iterative weight is

W_{ij} = diag$(w_1, w_2, \dots, w_k)$ and $w_{ij} = n_{i1}n_{i2}\mu_{ij}(1-\mu_{ij})$

And the working dependent variable is:

$$z_{ij} = \eta_{ij} + \left((\frac{y_{i1\pm}y_{i2\pm}}{n_{i1}n_{i2}} - \mu_{ij}) \frac{1}{\mu_{ij}((1-\mu_{ij})} \right) \quad (4)$$

The estimate of β_{ij} can be obtained from the Iterative Weighted Least Squares method (IWLS).The IWLS begins with an initial guess for the solution then uses evaluated estimates at the initial guess to come up with another estimate that is closer to the solution.

This process continues until it converges to the actual solution

At each iteration t, the coefficients are updated where

– $y_{ij\pm}$ is the response vector (containing 0's and 1's)

$-\, n_{ij}$is the ith subtotal

$-\mu_{ij}$is the vector of fitted response probabilities from the previous iteration, at each entry t the t-1 iteration is:

$$\mu_{ij,t-1} = \frac{1}{1+\exp(-x'_{i} b_{t-1})}$$

$W_{ij,t-1}$is a diagonal matrix, with diagonal entries $n_{ij}\mu_{ij,t-1}(1-\mu_{ij,t-1})$.

Results

The study population consists of people who visited the hospital. The response malaria is dichotomous coded 1 for patients whose malaria test are positive and 0 for negative and typhoid fever is also dichotomous coded 1 for patients who have typhoid and 0 otherwise. The explanatory variables sex and location are dichotomous and coded male-1, female-0 and rural-1,urban-0 respectively while the ages of the patients are categorized into 0.5< 4years, 4<9years, 9<15 years and 15 years and above.

Table 1: Parameter Estimates and Standard Error for the Marginal Logistic Model

Covariates	β_{i1}	s.e	odds	*p*-Value	β_{i2}	s.e	odds	*p*-Value
Constant	0.964	0.132	2.623	0.000	-0.623	0.124	0.505	0.000
Age	-0.755	0.050	0.470	0.000	0.263	0.044	1.301	0.000
Sex	-0.218	0.131	0.804	0.097	0.122	0.115	0.115	0.286
Location	0.409	0.168	1.505	0.015	0.261	0.152	1.299	0.085

Table 2: Parameter Estimates and Standard Error for the Joint Logistic Model

Covariates	β_{11}	Exp (β)	s.e	*p*-value	β_{10}	Exp (β)	s.e	*p*-value	β_{01}	Exp (β)	s.e	*p*-value	β_{00}	Exp (β)	s.e	*p*-value
Constant	-1.045	0.352	0.124	0.000	-0.118	0.889	0.011	0.001	-2.025	-2.025	0.134	0.000	-1.491	0.225	0.124	0.000
Age	-0.325	0.723	0.043	0.000	0.652	0.521	0.040	0.000	0.558	1.747	0.045	0.000	0.329	1.390	0.045	0.000
Sex	-0.031	0.970	0.013	0.012	-0.166	0.847	0.024	0.180	0.052	1.052	0.009	0.000	0.082	1.085	0.009	0.000
Location	0.503	1.654	0.221	0.023	0.120	1.127	0.025	0.000	0.101	1.106	0.017	0.000	-0.511	0.600	0.189	0.007
Model $\chi 2$				12.028	(0.0073)											

Table 3: Fitted Probability Distribution of Malaria and Typhoid by Age, Sex and Location.

Age	Sex	Location	Probability dist. of Malaria infection	Probability dist. of Typhoid fever infection
0.5<4	Female	Urban	0.724	0.336
		Rural	0.798	0.396
	Male	Urban	0.678	0.364
		Rural	0.761	0.426
4< 9	Female	Urban	0.552	0.397
		Rural	0.650	0.461
	Male	Urban	0.498	0.426
		Rural	0.599	0.491
9<15	Female	Urban	0.367	0.461
		Rural	0.466	0.526
	Male	Urban	0.318	0.491
		Rural	0.412	0.557
≥15	Female	Urban	0.214	0.527
		Rural	0.291	0.591
	Male	Urban	0.180	0.557
		Rural	0.248	0.620

Table 4: Joint Response Probabilities

	μ_{11}	μ_{10}	μ_{01}	μ_{00}
1	0.260	0.470	0.116	0.183
2	0.374	0.500	0.127	0.118
3	0 .254	0.429	0.122	0.196
4	0.361	0.459	0.133	0.128
5	0.203	0.316	0.187	0.238
6	0.296	0.203	0.343	0.158
7	0.198	0.282	0.195	0.253
8	0.289	0.307	0.212	0.169
9	0.155	0.194	0.287	0.303
10	0.233	0.214	0.309	0.207
11	0.151	0.170	0.298	0.320
12	0.227	0.187	0.320	0.221
13	0.117	0.112	0.413	0.377
14	0.180	0.124	0.438	0.266
15	0.114	0.096	0.426	0.396
16	0.175	0.107	0.451	0.282

Discussion

Table 1shows the estimated parameters and standard error. To test $H_0 : B_1 = 0$ in model 1, the age of a patient has significant ($p < 0.001$) impact on the probability of malaria infection, while controlling for sex and location. The estimated parameter value (-0.755) indicates that probability of malaria infection decrease as the age of patients increases. The variable sex has significant ($p<0.1$) impact on the probability of malaria infection, location also have a significant ($p < 0.05$) impact with patients from the rural having higher probability of infection. The odds for malaria infection for the male patients is 0.804 times the odds for malaria infection among the female patients while the odds for a patient from the rural is 1.505 times the odds for patient from the urban. Male patients are 1.131 times more likely to have typhoid fever than female patients. Patients from the rural are 1.299 times more likely to have typhoid fever than patients from the urban.

Age has a significant ($p< 0.001$) impact on the probability of typhoid fever infection ,the result shows that the probability of typhoid fever infection among the patient increases with age (0.263), location also have a significant ($p<0.1$)impact. Sex on the other hand has no significant ($p>0.05$) impact on the probability of typhoid fever infection among the study population.

Table 2 shows the estimated parameters and standard error (s.e) for the joint probability model. Age of a patient has a significant ($p< 0.001$) impact on the joint probability of malaria infection and typhoid in each model, while controlling for sex and location. Sex and location also have significant impact on the joint probability of malaria infection and typhoid ($p< 0.001$). Odds value greater than 1 indicate higher likelihood of co- infection while Odds value that are less than one indicates lower likelihood of co- infections For instance the odd of a male patient having malaria and typhoidis 0.970 Compared to the odds of a female patient. The odd of a patient from rural area having malaria and typhoid infection is 1.654 times the odds of a patient urban area.

Table 3 shows the response probabilities for malaria infection and typhoid fever infection. The result indicates that the probability of malaria infection decreases as the ages of patients' increases, for instance men who are from the rural have the probability of malaria infection to be 0.761, 0.599, 0.412 and 0.248 in age category 0.5< 4years, 4<9years, 9<15 years and 15 years and above respectively, the probability of infection is higher among patients from the rural irrespective of the sex. The probability of typhoid fever infection on the other hand increases as the age of the patient increases, female from the rural have the probability of typhoid fever infection to be 0.336, 0.397, 0.461 and 0.527 respectively in each of the age category

Table 4 shows the joint response probabilities of malaria and typhoid fever infection. Result from table 4 also indicates that the probability of a patient having malaria infection but not typhoid fever infection is higher among age category that are less than five (0.5<4) years than the other categories i.e.4<9 years, 9<15 years and 15 years and above.

Conclusion

This study provides evidence of the predictors which influence malaria infection, typhoid infection and malaria and typhoid co-infection among the study population.

The findings of this study indicate there is an association between malaria infection and predictors such age, sex and location, typhoid fever infection is related with age and location of the patients.

The test of the significance of the coefficients of the predictors showed that the predictors; location, sex and age are good predictors of malaria infection; age and location are good predictors of typhoid infection.

Also, it was found that the probability of malaria infection decrease as age increases while the probability of typhoid infection increases with age.

The test of dependence shows that malaria infection and typhoid infection are dependent in all age categories.

Recommendations

Although malaria infection occurs in all age group, the maximum prevalence of malaria occurs in childhood and adolescence. Hence individual within these groups (childhood and adolescence) should be given more attention to reduce infection rate among these groups.

Unhygienic practices that are associated with the spread of parasitic and bacterial infections should be avoided especially in rural areas as the study revealed that probability of infection is higher in these areas.

From the analysis, it was observed that malaria and typhoid infections are dependent, hence Nigeria government should provide health care centre by ensuring adequate stocks of anti-malarial drugs and vaccines for typhoid in these centre.

Further research could be done on three or more response variables and in prediction. This would even improve upon the scope.

References

Agomo C. O and Oyibo W. A. (2013). Factors Associated with Risk of Malaria Infection among Pregnant Women in Lagos, Nigeria *Infectious Diseases of poverty* 2:19

Birhanie, M., Tessema, B., Ferede, G., Endris, M., and Enawgaw, B. (2014) Malaria, Typhoid Fever, and their Co-infection among Febrile Patients at a Rural Health Center in Northwest Ethiopia:. Advances in Medicine, Article ID 531074. http://www.dx.doi.org, accessed 11thAugust ,2015

Francis E. G. C. (2010).History of the Discovery of the Malaria Parasites and their Vectors. *Cox Parasites & Vectors*. 3(5):1-9.

Glonek, G. F.V and McCullagh, P. (1995). Multivariate Logistic Models. *Journal of Royal Statistical Society.* 57 (3):533-546.

Hagenlocher M. and Castro, M. C. (2015) Mapping Malaria Risk and Vulnerability in the United Republic of Tanzania: A Spatial Explicit Model. *Population Health Metrics*.13:2

Igbeneghu, C., Olisekodiaka, M. J. and Onuegbu, J. A. (2009).Malaria And Typhoid Fever among Adult Patients Presenting with Fever in Ibadan, Southwest Nigeria. *International Journal of Tropical Medicine*.4 (3)112-115.

Igiri B. E, Ejezie C. G, Inyang-Etoh P. C., Jimoh O., Okoduwa SIR2, Otu-Bassey IB1 and Ogba OM(2018) Incidence of malaria and salmonellosis co-infection in Ahmadu Bello University Teaching Hospital, Zaria, North-West Nigeria, Clinical Sciences Research and Reports: 1(2): 2-5

Islam, M. A. Chowdhury, R. I. and Briollais l. (2012).A Bivariate Binary Model for Testing Dependence in Outcomes. *Bulletin of the Malaysian Mathematical sciences Society*. 35 (4): 845–858.

Islam, M. A., Alzaid, A. A., Chowdhury, R. I., and Sultan, K. S. (2013). A Generalized Bivariate Bernoulli Model with Covariate Dependence. *Journal of Applied Statistics*. 40(5): 1064–1075.

Jones M. M, Feng-Bin W, Naveen K.V., (2015) Modeling malaria and typhoid fever co-infection dynamics Mathematical Biosciences 264 :128–144

Kazembe, L. N., Kleinschmidt, I. and Sharp, B. L. (2006).Patterns of Malaria-related Hospital Admissions and Mortality among Malawian Children: An Example of Spatial Modelling of Hospital Register Data. *Malaria Journal.* 5 (93):1-10

Keong, B. C. M. and Sulaiman, W. (2006). Typhoid and Malaria Co-Infection–An Interesting Finding in the Investigation of a Tropical Fever. *Malaysian Journal of Medical Sciences.*13 (1):74-75.

Le Cessie, S and Van Houwelingen, J. C., (1994). Logistic regression for Correlated Binary Data. Applied Statistics. 43: 95–108.

Li, J and Wong, W. K. (2011). Two-Dimensional Toxic Dose and Multivariate Logistic Regression, with Application to Decompression Sickness. *Biostatistics Oxford Journals*. 12 (1):143-155.

Liang, K. Zeger, S. and Qaqish, B. (1992).Multivariate Regression Analyses for Categorical Data.*Journal of the Royal Statistical Society. Series B*. 54(1):3-40.

Lu, M. and Yang, W. (2012).Multivariate Logistic Regression Analysis of Complex Survey Data with Application to BRFSS Data. *Journal of Data Science.*10:157-173

McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models, 2nd edition. Chapman & Hall Publishers, London. 211pp.

Okafor, F. U. and Oko-Ose, J. N., (2012).Prevalence of Malaria Infections among Children Aged Six Months to Eleven years in a Tertiary Institution in Benin City, Nigeria. *Journal of Medicine and Medical Sciences*.1 (10): 273-279.

Pradhan, P. (2011).Co-infection of Typhoid and Malaria. *Journal of Medical Laboratory and Diagnosis.*2 (3):22-26.

Uneke, C. J. (2008). Concurrent Malaria and Typhoid Fever in the Tropics: The Diagnostic Challenges and Public Health Implications *Journal of Vector Borne Diseases.* 45:33–142.

Wang, F. (2011).Logistic Regression: Use & Interpretation of Odds Ratio (OR). Surveillance and Assessment Branch, AHW Community Health Science, the University of Calgary.2nd Edition.3pp

W.H.O. (2015a).An Operational Manual for Indoor Residual Spraying (IRS) for Malaria Transmission Control and Elimination. 2nd Edition. (7-9) pp.

Yee T. W. (2008). VGAM Family Functions for Bivariate Binomial Responses. Beta Version 0.7-6. http://www.stat.auckland.ac.nz/yee, accessed 29th August , 2015

Yee, T. W. and Dirnbock, T. (2009).Models for Analysing Species' Presence/Absence Data at Two Time Points. *Journal of Theoretical Biology*. 259: 684-694.

Yury, B., Sergey, C. , Konstantin S., Anatoly Z.(2019), A Mathematical Model of Regional Socio-Economic Development of the Russian Arctic Zone Resources 8: 45.