

Factors Affecting Knowledge & Consciousness of Bangladeshi People about Hepatitis B1 (HB1): An Application of Linear Logistic Regression

Md. Behzad Noor*, Rubaiyat Shaimom Chowdhury

Department of Business Administration, Shanto-Mariam University of Creative Technology, Dhaka, Bangladesh

*Corresponding author: bnoormphilru15@gmail.com

Received July 19, 2015; Revised July 24, 2015; Accepted July 27, 2015

Abstract This study introduces to identify the effects of various explanatory variables to the knowledge and consciousness about the Hepatitis B1 (HB1) applying one of the popular statistical analysis called Logistic Regression. More specifically, here the dependent variable for the logistic regression model has been considered as “HB1 Vaccination”, that is, the respondents who are more accustomed to take HB1 vaccination is regarded as the more conscious people in connection of keeping themselves away from getting infected by HB1 virus. Bangladesh Demographic and Health Survey (BDHS, 2007) has been used as the main sources of data. It has been reflected that the respondents who watch TV are 3.695 times more conscious than that of the respondents who don't watch TV. Respondents with high economic status are more aware than those of the respondents who have the less economical status. The respondents who get the protected drinking water are 1.6 times more likely to get the vaccination of HB1 than that of the reference category. Evidence shows that literate respondents are 2.69 times more aware in terms of getting HB1 vaccination than that of those illiterate respondents. So, the access to TV with having the higher economic status along with the strong educational background ensures the profound knowledge and consciousness about the pandemic.

Keywords: *hepatitis B, hepatitis B1, linear logistic regression model*

Cite This Article: Md. Behzad Noor, and Rubaiyat Shaimom Chowdhury, “Factors Affecting Knowledge & Consciousness of Bangladeshi People about Hepatitis B1 (HB1): An Application of Linear Logistic Regression.” *American Journal of Epidemiology and Infectious Disease*, vol. 3, no. 3 (2015): 70-75. doi: 10.12691/ajeid-3-3-4.

1. Introduction

Hepatitis B1 virus (HB1) infection is a worldwide problem and between 350 and 400 million persons is estimated to suffer from this infection. HBV infection is a contagious disease that may transmit vertically from mothers to their neonates or horizontally by means of blood products and body secretions. The first published report about HBV infection in Iran was in 1972 [8]. Hepatitis B is an infectious inflammatory illness of the liver caused by the Hepatitis B1 virus (HB1) that affects hominoidea, including humans. Originally known as "serum hepatitis" the disease has caused epidemics in parts of Asia and Africa, and it is endemic in China. About a third of the world population has been infected at one point in their lives including 350 million who are chronic carriers [2].

The virus is transmitted by exposure to infectious blood or body fluids such as semen and vaginal fluids, while viral DNA has been detected in the saliva, tears, and urine of chronic carriers. Perinatal infection is a major route of infection in endemic (mainly developing) countries [1]. Other risk factors for developing HB1 infection include working in a healthcare setting, transfusions, dialysis,

acupuncture, tattooing, extended overseas travel, and residence in an institution [5]. However, Hepatitis B1 viruses cannot be spread by holding hands, sharing eating utensils or drinking glasses, kissing, hugging, coughing, sneezing, or breastfeeding [2].

Hepatitis B1 virus is an hepadnavirus—*hepa* from *hepatotropic* (attracted to the liver) and *dna* because it is a DNA virus—and it has a circular genome of partially double-stranded DNA. The viruses replicate through an RNA intermediate form by reverse transcription, which in practice relates them to retroviruses [7]. Although replication takes place in the liver, the virus spreads to the blood where viral proteins and antibodies against them are found in infected people. The hepatitis B virus is 50 to 100 times more infectious than HIV [3].

The acute illness causes liver inflammation, vomiting, jaundice and, rarely, death. Chronic hepatitis B may eventually cause cirrhosis and liver cancer—a disease with poor response to all but a few current therapies. The infection is preventable by vaccination [1].

Hepatitis B virus (HBV) infection is a worldwide problem and between 350 and 400 million persons is estimated to suffer from this infection. HBV infection is a contagious disease that may transmit vertically from mothers to their neonates or horizontally by means of blood products and body secretions [4]. The first

published report about HBV infection in Iran was in 1972. In later studies, the rate of HBV infection was reported from 1% to 2.1% in 1977 while further reports stated higher rates (between 3.5% and 2.49 in both voluntary blood donors and general population from 1988 to 1993 [8].

Generally, it is estimated that about 1.5 to 2.5 million people are suffering from HBV infection in I.R. Iran, and some of them are carriers that may transmit infection to others unintentionally [6].

2. Sources of Data

The data of this study was taken from the 2007 Bangladesh Demographic and Health Survey (BDHS 2007). The BDHS 2007 is a nationally representative survey from 10,996 women age 15-49 and 3,771 men aged 15-54 from 10,400 household covering 361 sample points (cluster) throughout Bangladesh 134 urban areas and 227 in the rural areas. The data has collected from these six administrative divisions for the country- Barisal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet. The sample for the Bangladesh Demographic and Health Survey (BDHS) 2007 also covered the entire population residing in private dwellings units in the country. Administratively, Bangladesh was divided into six divisions, which in turn, each division were divided into zilas and upazilas. Each urban area in the upzila was divided into wards, and into mahallas within the ward; each rural area in the upazila was divided into union parishads (UP) and into mauzas within the Ups. This survey was based on a two-stage stratified sample of households. The urban areas were stratified into three groups, i) Standard metropolitan areas, ii) Municipality areas, and iii) Other urban areas. These divisions allowed the country as a whole to be easily separated into rural and urban areas. The 2007 BDHS sample was a stratified and multi stage cluster sample consisting of 361 primary sampling units (PSUs), 134 in the urban area and 227 in the rural area (PSUs). A total of 10,819 households, on average 30 households from each PSU, were selected for the sample using an equal probability systematic sampling technique, of which 10,461 were found to be occupied and 10,400 were successfully interviewed. Finally, the survey was designed to obtain 11,485 completed interviews with ever-married women age 10-49, covering 4,360 interviews from urban areas and 7,125 from rural areas. All ever-married women age 10-49 in selected households and ever-married men age 15-54 in every second households were considered as eligible respondents. But finally, a total of 11,178 eligible women age 15-49, 4,230 from urban areas and 6,948 from rural areas were selected in these households and 10,996, 4,151 from urban areas and 6,845 from rural areas were interviewed. Data for ever-married women age 10-14 have been removed from the data set to use for the present study. Accordingly 4,074 potential eligible men in every second households were selected, of them, 3,771 were successfully interviewed.

In this survey five questionnaires viz., household questionnaire, women's questionnaire, men's questionnaire, community questionnaire and facility questionnaire following MEASURE DHS Model Questionnaires have been used.

The survey was conducted to determine on the respondent's background characteristics (age, residential history, education, religion, media exposure etc.); reproductive history; knowledge about the different diseases (Hepatitis B1, HIV/AIDS, etc.) through the process of getting the information regarding the protective measures of the respondents, nutrition; vaccinations and health of children under age five; marriage; fertility preference; husband's background and respondent's work etc.

Data collected from field were edited, coded and processed at Mitra and Association using CSPro, a joint software product of the US Census Bureau, Macro International, and Serpro S.A.

The data has collected from these six administrative divisions for the country- Barisal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet. The present study utilizes the BDHS with having a sample of 3151 where 2000 are females and 1151 are males.

The present study utilizes the BDHS data, 2007 ever-married women of age 10-49 are considered by the study. Our study sample is 3151.

3. Analytical Methodology

3.1. Logistic Regression Analysis

An interesting method that does not require any distribution assumption concerning explanatory variables is Cox's linear logistic regression model (1972). The logistic regression model can be used not only to identify risk factors but also to predict the probability of success. The model is now widely used in research work to access the influence of various socio-economic and demographic characteristics for controlling the effect of other variables on the likelihood of the occurrence of the event of interest. There are a variety of multivariate statistical techniques that can be used predict a binary dependent variable from a set of independent variables. Multiple regression analysis and discriminate analysis are two related techniques but these techniques are applicable only when the dependent and independent variables are measured in interval scale under the assumption that they are normally distributed with equal variances. Linear discriminate analysis does not allow direct prediction of group membership, but the assumption of multivariate normality of the independent variables as well as equal variance-covariance matrices in the groups, is required for the prediction rule to be optimal. Logistic regression analysis is similar to a linear regression model where the dependent variable is a dichotomous one, coded as 1 (event occurring) and 0 (event does not occurring). The independent variables can be interval level or categorical; if categorical, they should be dummy or indicator coded. Let Y_i denote the dichotomous dependent variable for i th observation and $Y_i=1$, if i -th individual is a success (event occurs) and $Y_i=0$, if the i th individual is a failure (event does not occurs). Suppose that for each of the individuals k independent variables $X_{i1}, X_{i2}, \dots, X_{ik}$ are measured and it is assumed that Y_i 's are normally distributed with mean P_i and variance and P_i is defined as the probability of success; the logistic regression is of the form:

$$P_r(P_i) = P_r(Y_i = 1) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{-(\beta_0 + \beta_1 x_i)}} \quad (1)$$

Or equivalently,

$$P_r(P_i) = P_r(Y_i = 0) \frac{e^z}{1 + e^{-z}} \quad (2)$$

Where β_0 and β_1 are the regression coefficients estimated from the data; the model assumes the form:

$$P_r(P_i) = P_r(Y_i = 1) = \frac{e^z}{1 + e^{-z}} \quad (3)$$

Or equivalently,

$$P_r(P_i) = P_r(Y_i = 1) = \frac{e^z}{1 + e^{-z}} \quad (3(1))$$

Where, $z = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$.

From equation (3) and (4) completed; however, the logarithm of the ratio of P_i and $1 - P_i$ which is called logit of P_i turns out to be a simple linear function of X_{ij} .

We define,

$$\text{Logit}(P_i) = \text{Log} e \frac{P_i}{1 - P_i} = \sum_{j=0}^k \beta_j X_{ij} = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} \quad (4)$$

The logit is the logarithm of the odds of success, that is, the logarithm ratio of the probability of success to the probability of failure. It is also called the logistic transform of P_i and equation (4) is a linear logistic model. In a logistic regression, the parameters of the model are estimated using the maximum likelihood. The logistic model can be rewritten in terms of the odds of an event occurring. First, as P_i increases, so does logit (P_i) and second, logit (P_i) varies over the whole real line, whereas P_i is bounded only between 0 and 1. If P_i is less than 0.5, logit (P_i) is negative; and if P_i is greater than 0.5, logit (P_i) is positive. The equation can be written in terms of odds as:

$$\text{Odds} = \frac{P_i}{1 + P_i} = \exp \left(\sum_{j=0}^k \beta_j X_{ij} \right)$$

The exponential rise to the power β_j is the factor by which the odds change when j -th independent variables increase by one unit. If β_j is positive factor will be greater than 1, which means that the odds are increased; if β_j is negative factor will be less than 1, which means that the odds are decreased. When β_j is 0, the factor equal 1, which leaves the odds unchanged.

3.1.1. Measuring the Worth of the Model

There are various statistics that have been proposed for assessing the worth of the logistic regression model, analogous to those that are used in linear regression. We examined the two of the proposed statistics as follows:

3.1.2. R^2 in Logistic Regression Model

The worth of the linear regression model can be determined by using R-square, but R^2 computed as in linear regression should not be used in logistic regression, at least not when the possible values of Y are zero and one. It is

evident that R^2 can be dropped considerably for every miss fitted point, so, R^2 can be less than 0.9 even for near- perfect fitting. Cox and Wermuth (1992) also conclude that R^2 should not be used when Y has only two possible values, and show that frequently $R^2 \approx 0.1$ when good models are used.

Various alternative forms of R^2 have been proposed for binomial logit model.

Maddala (1983) proposed using

$$R^2 = 1 - \left\{ \frac{L(0)}{L(\hat{\beta})} \right\}^{\frac{2}{n}} \quad (5)$$

With $L(0)$ denoting the likelihood for the null model (i.e., with no regressors) and $\left(L(\hat{\beta}) \right)$ representing the likelihood function that would result when replaces in the following equation

$$q(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n P_i^{Y_i} (1 - P_i)^{1 - Y_i} \quad (6)$$

Essentially the same expression, except that $2/n$ was misprinted as $1/n$, was given by Cox and Snell (1989). [Equation (5) is motivated by the form of the likelihood ratio test for testing the fitted model against the null model.

It can be shown that R^2 as defined in the linear equation is equivalent to the right hand side of the equation (5). Hence, this is a natural form of the R^2 in the logistic

regression.] Since, the likelihood function $\left(L(\hat{\beta}) \right)$ is a

product of probabilities, it follows that the value of the function must be less than 1. Thus, the maximum possible value for R-square defined by equation 5) is

$\max R^2 = 1 - \left\{ \frac{L(0)}{L(\hat{\beta})} \right\}^{\frac{2}{n}}$. In linear regression model $\hat{Y} - \bar{Y}$ is

used for the null model. Similarly, in logistic regression we would have $P - \gamma$ for the null model, with γ_1 denoting the percentage of the 1's in the data set. It follows that

$\max R^2 = 1 - \left\{ \gamma_1^{\gamma_1 n} (1 - \gamma_1)^{1 - \gamma_1 n} \right\}^{\frac{2}{n}}$. For example, if $\gamma_1 = .5$,

then $\max R^2 = .75$. This is the largest possible value of the R^2 defined by equation (5). When the data are quite sparse, the maximum possible value will be close to zero.

Therefore, Nagelkerke (1991) suggests that \bar{R}^2 be used, with $\bar{R}^2 = R^2 / \max R^2$.

3.1.3 Correct Classification Rate (CCR)

We may criticize any statistics that is a function of the \hat{P}_i when Y is binary. Each \hat{P}_i and its closeness to Y_i depends on more than the worth of the model. If our objective is to predict whether a subject will or will not have the attribute of interest, a more meaningful measure of the worth of the model would be the percentage of the subjects in the data set that classify correctly. Accordingly, we will use the correct classification rate (CCR) as a measure of the fit of the model.

3.2. Variable Selection in the Model

To apply the logistic regression model, we need to re-code of explanatory variables. For the sake of making our

analysis more reliable and understandable it is indispensable to get an idea about the coding of the selected predictor variables. The re-coding of the selected dependent and independent variables are mentioned in the following table:

Table 1. List of dependent and Explanatory Variables with Category for Logistic Regression Model

Dependent Variables (Y)	Explanatory variables (X)	Model No.
<p>2. Hepatitis B1 Vaccination</p> <p>Category: (Take HB1 Vaccination = 1 and Don't take HB1 Vaccination=0)</p>	<p>a) Age of the diabetes patients. Category: Age (Category: ≤20 = 1, 21-30 =2, 31-40 = 3 and 41-50 = 4</p> <p>b) Educational Status Category: Literate = 1 and Illiterate = 0</p> <p>d) Occupation of the Respondents Category: In Job = 1, Business = 2, Agriculture = 3, Labour = 4 and Others = 5</p> <p>e) Watching TV Category: Yes = 1 and No = 0</p> <p>f) Economic Status Category: High Economic Status = 1 and Low Economic Status = 0</p> <p>g) Place of Residence Category: City =1and Village= 0</p> <p>h) Sources of Drinking Water Category: Unpolluted = 1 and Polluted = 0</p> <p>i) Access to Newspaper Category: Yes = 1 and No = 0</p>	Model No.-1

4. Result and Discussion

Table 2 represents the differential patterns of the different variables to the selected dependent variables titled "Hepatitis B1 Vaccination" with the selected odds ratio.

Here from the table it has been found that the age categories of 31-40 and 41-50 contain more odds ratio (i.e.,

3.56 and 2.56 respectively) than the other categories. So, those two categories' respondents are more used to use hepatitis B vaccination than the reference categories. From this finding it could be interpreted that the middle aged population are more conscious than the categorical respondents belong to the age category of <20 and 21-30 years in the case of having the HB1 vaccination (Table 2).

Table 2. Logistic Regression for the Effects of Selected Independent Variables on Reducing Chance of the Prevalence of Hepatitis B1 through HB1 Vaccination

Characteristics	Coefficient (β)	S.E of estimates (β)	ρ Values	Relative risk (Odds ratio)
Age				
<20	-0.365	0.561	0.032***	0.563
21-30 (RC)	1.000
31-40	0.563	0.698	0.045*	3.561
41-50	0.653	0.2365	0.0236	2.563
Educational Status				
Illiterate (RC)	1.000
Literate	0.556	0.298	0.165	2.695
Occupation				
In Job	1.000
Business	-0.256	0.563	0.035	0.632
Agriculture	-0.563	0.523	0.025	0.523
Labour	-0.254	0.563	0.025	0.563
Others	-0.563	0.653	0.0256	0.356
Watching TV				
NO (RC)	1.000
Yes	0.683	0.217	0.002***	3.695
Economic Status				
High Economic Status (RC)	1.000
Low Economic Status	-0.365	0.385	0.405	0.556
Place of Living				
Urban (RC)	1.000
Rural	-0.593	0.456	0.039**	0.809
Sources of Drinking Water				
Unprotected (RC)	1.000
Protected	0.365	0.456	0.168	1.653
Access to Newspaper				
No (RC)	1.000
Yes	0.457	0.412	0.236	1.657
Constant	-1.533	0.348	0.000	0.216
-2 Log likelihood = 720.622				
R Square = 0.76				

Evidence has been found that watching TV provides more information regarding any aspects which has a greater impact on the storage of knowledge. Here from [Table 2](#) the respondents who have the more access to TV have the more level of consciousness as they are 3.695 times more habituated to take HB1 vaccination than that of the respondents who don't have the access to TV. This variable has also the significant contribution to the respective dependent variable ([Table 2](#)).

Urban people are more facilitate to have several types of information from different sources than of those people who belong to rural place. It has been amplified that rural people are 0.809 times more compact to have the latest information in connection of being conscious by taking HB1 vaccine to get away from the pandemic of hepatitis B1 than of those who lives in the urban place ([Table 2](#)).

The people who used to take the unpolluted water have the more tendency of acquiring HB1 vaccination than of those who are used to take the polluted water ([Table 2](#)).

4.1. R^2 in the Logistic Regression

For the above fitted model the Cox and Snell $R^2 = 0.76$ and the Negelkereke $\bar{R}^2 = 0.65$. It is observe that when the value of \bar{R}^2 is exceeds 0.5 the data fit the binary logistics regression model well. Therefore the model can be used for prediction of the significant effect of selected independent variables on reducing the chances of the prevalence of Hepatitis B1 (HB1) by adopting HB1 vaccine.

4.2. Correct Classification Rate

Furthermore we will use the Correct Classification Rate CCR as a measure of the fit of the model. In order to find the CCR we have the following tables.

If we use 0.5 cut as the threshold or cut value, we have from [Table 4](#), CCR = 0.87. Since a model that affords better classification performance should be judged superior by a goodness-of-fit test that indirectly assesses the classification performance of the model. Through classification performance we conclude that our fitted model may be used for prediction.

Table 3. Observed Classification Table ^{a, b}

Habit of Taking HB Vaccination		Predicted		Percentage Correct
		No	Yes	
Observed	No	720	21	95
	Yes	277	10	5.0
	Overall Percentage			84.3

a. Constant is included in the model

b. The cut value is 0.5

Table 4. Predicted Classification Table

Habit of Taking HB Vaccination		Predicted		Percentage Correct
		No	Yes	
Observed	No	867	0	100
	Yes	123	7	5.4
	Overall Percentage			86.6

a. Constant is included in the model

b. The cut value is 0.5

5. Conclusion

It has been found from the logistic regression that watching TV provides more information regarding any aspects which has a greater impact on the storage of knowledge. Here from [Table 1](#) the respondents who have the more access to TV have the more level of consciousness (3.695 times) than that of the respondents who don't have the access to TV. Higher economic status holders have more awareness than those of the respondents who have the less economical status. This evidence can be interpreted as the higher economical status holder has the more access to verities informative equipment that makes the respondents more conscious. Sources of drinking water has the contribution to the selected dependent variable as we have got that the respondents who get the protected drinking water are 1.6 times more likely to get the vaccination of HB than that of the reference category.

Although hepatitis B is of immense public health importance in the WPR, the motivation for setting a regional hepatitis B control goal goes beyond the control of hepatitis B [5]. The proposed hepatitis B control goal provides an outcome indicator for monitoring both the quantity (i.e. coverage) and quality of routine immunization services. This prompted a review of immunization quality, including the timing of the birth dose and the incidence of vaccine freezing, which may render the vaccine impotent [5]. Most importantly, more comprehensive research is needed to the better understanding of the factors relevant to the development of risk reduction interventions for the prevention of the diseases and to let the people know about the risk factors those are associated with this disaster pandemic, so that they could be aware of those risky factors.

Acknowledgement

At the outset all praises and utmost gratitude to Almighty Allah, source of all power and knowledge, for giving me strength, patience and ability to accomplish my research work. This research paper is a part of my research work at M. Phil. level whole heartedly I have tried my best to bring out good research work with all my sincerity, honesty, merit and hard labor.

This research work has been completed and made possible only through the sympathetic help enthusiastic assistance and proper guidance, given so logically and systematically by honorable supervisor Dr. J.A.M Shoquilor Rahman, Professor, Department of Population Science and Human Resource Development, University of Rajshahi and co-supervisor Dr. Nazrul Hoque, Ph.D. Associate, Director for, Estimate and Projections, Faculty Research Associate, the University of Texas at San-Antonio Texas.

Finally, I would like to thank my little sweet daughter for creating the source of inspiration. Lastly, I am grateful to my family members, especially to my beloved grandfather & parents for their continuous inspiration throughout the whole period of my study life.

References

- [1] Centers for Disease Control and Prevention, *Kaposi's Sarcoma and Pneumocystis Pneumonia among Homosexual Men, 2011* NY City and California *MM Weekly Report*.
- [2] Centers for Disease Control and Prevention, *West Nile Virus*, Accessed on September 8, 2010, Available: <http://www.cdc.gov/ncidod/dvbid/westnile/index.htm>.
- [3] Duke University Medical Center Library: *Conducting systematic reviews*, 2012, Available: <http://guides.mclibrary.duke.edu/sysreview>.
- [4] Garson, G. D., *Guide to writing empirical papers, theses, and dissertations*, 2002, NY Marcel Dekker.
- [5] Hewitt, J, ENGI, *Template for taking research note*, 2012, Retrieved from Rice Center for Engineering, Available: <http://rcel.rice.edu/engi600/> Thesis Literature Review Handbook v.1 53.
- [6] Primary health care. (n.d.), *In MeSH database*, 2011, Available: <http://www.ncbi.nlm.nih.gov/mesh?term=primary%20health%20care>.
- [7] Ungchusak, K., P. Auewarakul, S.F. Dowell, et al., *Probable Person-to-Person Transmission of Avian Influenza A (H5N1)*, *New England Journal of Medicine*, 2005, 352:333-40.
- [8] Weiss, R.A. and McMichael, A.J., *Social and environmental risk factors in the emergence of infectious diseases*, *Nature Medicine*, 2010, 10, S70-S76.