© 2015 IJSRSET | Volume 1 | Issue 4 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Science

Determining Factors in Reducing the Chances of the Prevalence of HIV/AIDS by Using Condom: A Study on Ever-Married Bangladeshi Population

Md. Behzad Noor, Rubaiyat Shaimom Chowdhury
Department of Business Administration, Shanto-Mariam University of Creative Technology, Dhaka, Bangladesh

ABSTRACT

This study introduces to identify the effects of various explanatory variables to the knowledge and consciousness about the HIV/AIDS applying one of the popular statistical analysis called Logistic Regression. Bangladesh Demographic and Health Survey (BDHS, 2007) has been used as the main sources of data. It has been reflected that the age category of 31-40 and 41-50 are 1.36 and 2.56 times more habituated to use condom than the reference category. So, it is the age group of <20 years of age who are more exposed to the risk of getting the HIV/AIDS than the other categories as they are less likely to use condom as the preventive measure. It is amplified that the respondents who responded as they used to watch TV at a regular basis are 3.69 times more conscious about the prevalence procedure of HIV/AIDS than those who don't have the habit of watching TV. Urban people are more aware in terms of having more habit of using condom for keeping themselves way from HIV/AIDS that of the counter part of the respondents who belong to some rural place (0.809 times more likely to use condom than the reference category).

Keywords: HIV/AIDS, Logistic Regression Model, Bangladesh, Population.

I. INTRODUCTION

Knowledge and consciousness are the most important factors in gaining different diseases oriented information to keep one-self being aware from the pandemic like HIV/AIDS. This study has been introduced to have some idea regarding the factors affecting the knowledge and consciousness about the most intensified diseases (HIV/AIDS) and to discover the probable way of maintain the safeguard.

HIV and AIDS have changed the way we go about our daily lives. We have seen an increase in the prevention, treatment and care for this disease over the years, at the same time more lives have been lost to this preventable and fatal disease. In 2008, there were about 32.2 million People Living with HIV (PLHIV) worldwide,1 2.7 million people newly infected with the virus annually and about 2 million deaths. 1 In the Pacific region, there were about 29,629 reported HIV cases and 5,162 new cases in 2008 (Ontario Public Health Library Association (OPHLA), (2008)). Although most Pacific Island Countries (PIC) are classified as low prevalence, the incidence of HIV has been increasing since the first case was reported in 1984 (Leenaars *et al.,* (2012)).

When someone has HIV, it can be present in potentially infectious quantities in blood, semen, vaginal fluids, rectal secretions and breast milk. Transmission of HIV can happen if one of these fluids gets into somebody else's body, either directly into the bloodstream (such as in injecting drug use, or from an HIV-positive woman to her baby), or through unprotected anal, or vaginal sex, and much less often, unprotected oral sex (Taylor, (2010)).

Various factors affect the risk of HIV transmission. How likely it is that transmission will occur is directly linked to the viral load of the HIV-positive person. The more virus that is present, the more likely it is to be passed on.

In this context we sometimes talk about ‘infectiousness’ (Gomez-Gonzalo et al., (2012)).

The World Health Organization (WHO) collects information on global deaths by International Classification of Diseases (ICD) code categories. The following table lists the top infectious diseases killers which caused more than 100,000 deaths in 2002 (estimated). 1993 data is included for comparison (Weiss *et al.,* (2010)).

Table 1 Ranking of the to the Infectious Diseases Killer in the Global Area

Rank	Cause of Death	Death 2009	Percentage of all Deaths	Deaths 1993	1993 Rank
N/A	All Infectious Diseases	14.7 million	25.9%	16.4 million	32.2%
1	Lower respiratory infections	3.9 million	6.9%	4.1 million	1
2	HIV/AIDS	2.8 million	4.9%	0.7 million	7
3	Diarrheal diseases	108 million	3.2%	30 million	2
4	Tuberculosis (TB)	1.6 million	2.7%	2.7 million	3
5	Malaria	1.3 million	2.2%	2.0 million	4
6	Measles	0.6 million	1.1%	1.1 million	5
7	Pertussis	0.29 million	0.5%	0.36 million	7
8	Tetanus	0.21 million	0.4%	0.15 million	12
9	Meningitis	0.17 million	0.3%	0.25 million	8
10	Syphilis	0.16 million	0.3%	0.19 million	11
11	Hepatitis B	0.10 million	0.2%	0.93 million	6
12-17	Tropical diseases	0.13 million	0.2%	0.53 million	9, 10, 16-18

Source: WHO, 2009

The top three single agent/disease killers are HIV/AIDS, TB, and Malaria. While the number of deaths due to nearly every disease have decreased have decreased, deaths due to HIV/AIDS have increased fourfold. Childhood diseases include pertussus, poliomyelitis, diphtheria, measles and tetanus. Children also make up a large percentage of lower respiratory and diarrheal deaths (Weiss *et al.,* (2010)).

It is notified that HIV/AIDS and HB as appeared as two silent killer diseases and become threatened to modern civilization for developed and under developed countries. So any study regarding knowledge and consciousness is very important to have the clear idea regarding this issue. In contrast, Human Immunodeficiency Virus (HIV) kills its victims very slowly by attacking their immune system (World News, (2012)). As a result, a lot of affected persons transmit the virus to many others before even realizing that they are carrying the diseases. Also, the relatively low virulence always its victims to travel along distance, increasing the likelihood of an epidemic (World Health Organization, (2008)).

II. METHODS AND MATERIAL

2.1 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 aged15-54 from 10,400 household covering 361 sample points (cluster) throughout Bangladesh 134 urban areas and 227 in the rural areas. The daya has collected from these six administrative divisions for the country-Barisal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet. 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.

2.1.1 Data Collection Procedure

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 and 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 vize., households 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 and use of family planning methods; antenatal and delivery care; 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 easiest procedure of analyzing the data is to use computer program. At present nobody thinks to analyze data without a suitable computer program. No other alternative is available to analyze the data quickly, easily and correctly. The SPSS' 2007 computer program has been used for analyzing the data.

2.2. Analytical Methodology

2.2.1 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 ices 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 ith observation and $Y_i=1$, if i-th individual is a success (event occurs) and $Y_i=0$, if the ith individual is a failure (event does not occurs). Suppose that for each of the individuals k independent variables $X_{i1}, X_{i2}, ...X_{ik}$ are measured and it is assumed that Y_i's are normally distributed with mean P_i and variance and Pi is defined as the probability of success; the logistic regression is of the form:

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

Or equivalently,

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

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

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

Or equivalently,

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

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

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,

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

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.

1.2.1.1 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

2.2.1.2 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

$$R^2 = 1 - \left\{\frac{L(0)}{L(\beta)}\right\}^{\frac{2}{n}} \text{.................................(1)}$$

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

$$q(Y_1, Y_2, ..., Y_n) = \prod_{i=1}^{n} P_i^{Y_i} (1-P_i)^{1-Y_i} \text{ ......(2)}$$

Essentially the same expression, except that 2/n was misprinted as 1/n, was given by Cox and Snell (1989). [Equation (2) 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 (2). 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 (2) is max. $R^2 = 1 - \{(L0)\}^{\frac{2}{n}}$. In linear regression model $\hat{Y} - \overline{Y}$ is used for the null model. Similarly, in logistic regression we would have P-γ for the null model, with γ_1 denoting the percentage of the 1's in the data set. It follows that max $R^2 = 1 - \left\{ \gamma_1^{\gamma_{1n}} (1-\gamma_1)^{1-\gamma_{1n}} \right\}^{\frac{2}{n}}$. For example, if γ_1= .5, then max R2 = .75. This is the largest possible value of the R2 defined by equation (2). When the data are quite sparse, the maximum possible value will be close to zero. Therefore, Nagelkereke (1991) suggests that $\overline{R}^2$ be used, with $\overline{R}^2 - R^2 / \max R^2$.

2.2.1.3 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.

2.2.2 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. For the availability of data we consider two linear logistic regression models. To fit the logistic regression model we select most important explanatory variables which are significantly associated with the dependent variables that were shown in chapter three. The dependent variable for model is Use Condom at the time of Sexual Intercourse. The categories of the dependent and explanatory variables are mentioned in Table 2.

Table 2.2.2 List of Dependent and Explanatory Variables with Category for Logistic Regression Model

Model No.	Dependent Variables (Y)	Explanatory variables (X)
Model No.- 1	Use Condom at the Time of Sexual Intercourse (For model No.-1) Category: (Yes = 1 and No =0)	a) Age of the diabetes patients. Category: Age (Category: ≤20 = 1, 21-30 =2, 31-40 = 3 and 41-50 = 4 b) Educational Status Category: Literate = 1and Illiterate = 0 c) Occupation of the Respondents Category: In Job = 1, Business = 2, Agriculture = 3, Labour = 4 and Others = 5 d) Watching TV Category: Yes = 1 and No = 0 e) Economic Status Category: High Economic Status = 1 and Low Economic Status = 0 f) Place of Residence Category: City =1and Village= 0 g) Sources of Drinking Water Category: Unpolluted = 1 and Polluted = 0 h) Access to Newspaper Category: Yes = 1 and No = 0

III. RESULTS AND DISCUSSION

The following Table 3 represents the effects of various explanatory variables to our first dependent variable "Use Condom at the Time of Sexual Intercourse" where, regression co-efficient with the corresponding significance level and the odds ratios are revealed.

Table 3 : Logistic Regression for the Effects of Selected Independent Variables on Reducing Chance of HIV/AIDS by Using Condom at the Time of Sexual Intercourse

Characteristics	**Coefficient (β)**	**S.E of estimates (β)**	ρ Values	Relative risk (Odd ratio)
Age				
<20	-0.689	0.384	0.073*	0.36
21-30 (RC)	……… …	………… …….	……… …	1.000
31-40	0.517	0.589	0.256	1.36
41-50	0.317	0.267	0.235	2.56
Educational Status				
Illiterate (RC)	....	....	....	1.000
Literate	0.336	0.216	0.120	3.695
Occupation				
In Job	……… ….	………… …….	……… ….	1.000
Business	0.567	0.226	0.008**	0.85
Agriculture	-0.236	0.56	0.589	0.35
Labour	-0.390	0.423	0.356	0.2
Others	-0.52	0.213	0.036***	0.52
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	-0.258	0.310	0.405	0.773
Place of Living				
Urban (RC)	....	....	....	1.000
Rural	-0.593	0.271	0.029**	0.809
Sources of Drinking Water				
Unprotected (RC)	....	....	....	1.000
Protected	0.286	0.208	0.168	1.332
Access to Newspaper				
No (RC)	....	....	....	1.000
Yes	0.375	0.378	0.322	1.454
Constant	-1.533	0.348	0.000	0.216
-2 Log likelihood = 720.622 **R Square = 0.**80				

Here the age category of 21-30 is considered as the reference category. With respect to the reference category it can be noticed that the respondents of the age category of 31-40 and 41-50 are 1.36 and 2.56 times more habituated to use condom than the reference category. So, it is the age group of <20 years of age who are more exposed to the risk of getting the HIV/AIDS than the other categories as they are less likely to use condom as the preventive measure (Table 3).

As far as the occupational category is concerned it can be stated that the respondents who are the job holders who are more frequent to use condom than other categories as the odds ratio of the other occupational categories are 0.85, 0.35, 0.2 and 0.52 for the occupational categories of business, agriculture, labour and others respectively. Here among the occupational categories the business group and the others group have the significant impact on the respective dependent variable (Table.3).

Contribution of the mass media in accessing knowledge and awareness of the people is undoubtedly extraordinary. Here also from Table 3 it is amplified that the respondents who responded as they used to watch TV at a regular basis are 3.69 times more conscious about the prevalence procedure of HIV/AIDS than those who don't have the habit of watching TV in connection of gathering varieties of information to be conscious. It has also the significant effect on the respective dependent variable (Table.3).

Urban is the place where people used to get more facilities through availing those they become more aware. From Table.3.1 it has been detected that the urban people are more aware in terms of having more habit of using condom for keeping themselves way from HIV/AIDS that of the counter part of the respondents who belong to some rural place (0.809 times more likely to use condom than the reference category).

3.1 R^2 in the Logistic Regression

For the above fitted model the Cox and Snell $R^2 = 0.80$ and the Negelkereke $\overline{R}^2 = 0.70$. It is observe that when the value of $\overline{R}^2$ is exceeds 0.5 the data fit the binary logistics regression model well. Therefore the model can be used for prediction of significant effect of selected independent variables on reducing chance of HIV/AIDS by using condom at the time of sexual intercourse.

1.3 Correct Classification Rate (CCR)

Further we will use the correct classification rate (CCR) as a measure of the fit of the logistic regression 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 3.2, CCR = 0.88. 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 3.2 Observed Classification Table $_{a,\,b}$

Use Condom at the time of Sexual Intercourse		Predicted		Percentage Correct
		No	Yes	
Observed	No	713	7	95
	Yes	264	13	5.0
	Overall Percentage			87.0

Table 3.2.1 Predicted Classification Table $_{a}$

Use Condom at the time of Sexual Intercourse		Predicted		Percentage Correct
		No	Yes	
Observed	No	867	0	100
	Yes	123	7	5.4
	Overall Percentage			87.7

IV. CONCLUSION

It has been found from the logistic regression that the respondents of the age group 41-50 are 2.56 times more conscious than the reference category (i.e., age group of 21-30 years.). So far as the occupational category is concerned it has been got that with respect to the job holders the rest of the occupational categories have less odds ratio (i.e., for business, agriculture, labour and others the odds ratios are 0.85, 0.35 0.2 and 0.52 respectively) in terms of the habit of using condom at the time of sexual intercourse. Again the literate persons are found 3.695 times more likely to use protective measure than that of illiterate respondents. Here also it is amplified that the respondents who responded as they used to watch TV at a regular basis are 3.69 times more conscious. Economic status has a very indispensable effect on the selected dependent variable as it is seen that the respondents holding the higher economic status has the more likelihood to use condom at the time of intercourse than of those who have comparatively lower economical status (0.77 times more likely to use condom as the preventive measure). Our research also provides the output as the respondents who have the attachment with the daily newspaper are 1.45 times more conscious of the prevalence of HIV/AIDS than of those who don't have the access to the newspaper.

Specifically it could be stated that the opportunity should be created to get the data on the hepatitis b and HIV/AIDS affected people directly to get closer about the sources of infection and by exploring those data the core research could be made to detect factors affecting the prevalence of the pandemic.

V. REFERENCES

[1]. Gomez-Gonzalo, K.L. 2012. Factors in the Emergence of Infectious Diseases. Emerging Infectious Diseases 1:7–15.

[2]. Leenaars, M., Hooijmans, C. R., van Veggel, N., ter Riet, G., Leeflang, M., Hooft, L., Ritskings-Hoitinga, M. 2012. A step-by-step guide to systematically identify all relevant animal studies, Laboratory Animals, 46(1), 24–31.

[3]. Taylor, D. 2010. The literature review: a few tips on conducting Retrieved from the University of Toronto, Health Sciences Writing Centre, website at http://www.writing.utoronto.ca/advice/specific-types-of writing/literature-review

[4]. Weiss, R.A. and McMichael, A.J. 2010. Social and environmental risk factors in the emergence of infectious diseases, Nature Medicine, 10, S70–S76.

[5]. World Health Organization. 2010. Progress on Global Access to HIV Antiretroviral Therapy: An Update on 3 by 5, September 8, 2010. http://www.who.int/3by5/fullreportJune2005.pdf.

[6]. World News. (2012). Literature review [Video]. Retrieved from http://wn.com/literature_review