Logistic Regression Analysis of Syphilis in Birnin Kebbi

Babayemi¹ A. W. and Olorunpomi² O. T
¹Department of Mathematics, Kebbi State University of Science and Technology, Aliero, Kebbi State, Nigeria
²Department of Mathematics, Nigeria Police Academy; Wudil, Kano State, Nigeria

Abstract—This work examined the prevalence, demographic and estimated factors associated with people living with syphilis in Birnin Kebbi. Data were collected through questionnaires and administered at random in four different areas in Birnin Kebbi. These areas include Sir, Yahaya Memorial hospital, Federal Medical Centre, Godiya Memorial Hospital and Zauro Memorial Hospital. Method of Logistic Regression was implemented for the analysis. The result of the study revealed that the odd ratio of the respondents with symptoms such as sore throat, swollen lymph and rashes on the palm of their hands is 5.12, meaning that the chances or odds of contacting syphilis are 5.12 times higher for these respondents than ordinary person without these symptoms.

Index Terms—Syphilis, Demographic and estimated factors, Logistic Regression and Odd Ratio.

I. INTRODUCTION

Syphilis is a sexually transmitted disease caused by a bacterial infection of the bacterium treponema pallidum. Syphilis can be passed from one person to another during sexual contact that involves vaginal, oral, or anal sex. Syphilis infection can also be passed from an infected mother to her baby during pregnancy. Until the 1970s, the disease was endemic to many parts of the world, including the Middle East; aggressive treatment programs abated its prevalence, but such programs have since ceased. Transmission occurs through contact with infectious lesions on the skin and mucous membranes and with contaminated drinking vessels. Walker (2002), in his study focus on the history of the diseases, most involve late stage syphilis because the neurological damage common to late stage syphilis provides an excuse for strange behaviors. Schmudinn and Hoffman (1998), discovered and isolated the bacterium that causes syphilis. In (1906), German bacteriologist August Von Wassermann, working in conjunction with Albert Neisser, discovered the Wassermann reaction, a blood-serum test that could determine if a person had syphilis. German scientist Ehrlich, in (1908), began his research to find a better drug to fight the disease by testing hundreds of different arsenic compounds on syphilitic rats. One compound was found that effectively destroyed syphilis without destroying the rat. He called it Salvarsan, which in English means “I save”. Treponema palladium is the bacterium that causes syphilis. Syphilis can move throughout the body, damaging many organs over time. After initial penetration, the bacteria enter the lymph capillaries, where they are transported to the nearest lymph gland. There, they multiply and are released into the bloodstream, where they invade every part of the body (Greenblatt, 1988; Gordin, 1990; Gerbase, 1998; Garg, 2001; Golden, 2003; Gunn, 2003; Gupta, 2004).

Holmes (1999), the worldwide re-emergence of syphilis brought this complex sexual infection back to mainstream medicine. Since the late 1990s there has been a well documented resurgence of infectious syphilis beginning in Europe, predominately among men who have sex with men (MSM) syphilis has a myriad of presentations and can mimic many other infections and immune-mediated diseases. The complex and variable manifestations of the disease mean that vigilance is required in every medical discipline. Reports of local outbreaks have made reference to the diagnostic difficulties in order to remind clinicians to think of syphilis when encountering such patients, the control of syphilis requires early identification and treatment of cases. This calls for tests that are easily administered and interpreted, and treatment that is fast, efficacious and side effect free. Stemming the spread of infection will also require increased education and prompt notification of partners. Syphilis is a sexually transmitted disease that begins with genital sores, progresses to a general rash, and then to disfiguring abscesses and scabs all over the body. In its late stages, untreated syphilis can cause heart abnormalities, mental disorders, blindness, other neurological problems, and death. It appeared prominently in Europe at the end of the fourteenth century, and by 1500 syphilis had spread to the other continents. Syphilis remains as an important cause of morbidity, mortality and a possible transmission factor in the spread of HIV infection. With the increasing prevalence of infectious syphilis clinical vigilance and increased testing for syphilis is warranted. The quality of serological diagnosis is improved by
using a combination of treponemal and non treponemal tests; this is common practice after a positive screening result is obtained. Further development of point of care screening kits currently underway will aid with expansion of screening programs. Effective treatment is widely available in the form of penicillin. However the therapeutic options for those patients allergic to penicillin are limited. There is a need for further evaluation in large scale randomized controlled trials of treatment regimens, in particular regimen dose and duration and penicillin alternatives (Collee, 1989; Surveill, 1999; Choy, 2003; John , 2008; Choudhry, 2010).

A serological screening was carried out during the period of August—October, 2011 to assess the risk of infection with syphilis and co-infection with HIV-1/2, HBV, and HCV among these attendees. Unlinked and coded serum samples received from 417 subjects (260 females and 157 males) was screened by laboratory tests commonly used for laboratory diagnosis of HIV, HBV, HCV, and syphilis. Among the 417 samples serological reactivity was detected for HIV-1/2 in 27(6.5%), HBV in 15(3.6%), HCV in 4(1.0%), and syphilis in none (0.0%). The incidence of HIV-1/2, HBV and HCV was higher among males than females, i.e. 18/157 (11.5%) versus 28/260 (10.8%). None was found to have co-infection with HIV-HBV, HBV—HCV, HIV—syphilis, HIV—HCV, HBV—syphilis, and HCV—syphilis. Age, sex, marital status, history of vaccinations, and locality significantly influence the rate of HIV, HBV and HCV positivity (Okonko, 2013).

A. AIM AND OBJECTIVES OF THE STUDY
This study aimed at examining the prevalence, socio-demographic, behavioral and institutional factors associated with syphilis in Birnin Kebbi. The objectives are:
i. To give an insight for the prevalence of syphilis infection.
ii. To estimate the number of people living with syphilis.
iii. To predict the odd of having the disease.

B. SIGNIFICANCE OF THE STUDY
i. Add to the existing literature on the syphilis infection.
ii. give the government health practitioner, researcher insight on how to control the spread of the disease.

C. SCOPE AND LIMITATION
The work examines the prevalence and factors associated with Syphilis in Birnin Kebbi. The method to be used in this research work is logistic regression methodology, the data are collected through questionnaire, and this involves four hospitals. Sir Yahaya Memorial Hospital, Godiya Memorial Hospital, Federal medical centre, and Fati lami memorial hospital.

II. METHOD OF DATA ANALYSIS
Logistic regression: is a type of regression analysis used for predicting the outcome of a categorical dependent variable (a dependent variable that can take on a limited number of categories) base on one or more predictor variables. The probabilities describing the possible outcome of a single trial are modeled, as a function of explanatory variable, using a logistic function. Logistic regression measures the relationship between a categorical dependent variable and usually a continuous independent variable (or several), by converting the dependent variable to probability scores. As such it treats the same set of problems as probit regression using similar techniques. Logistic regression Analysis is used to investigate relationship. It is also use to establish an equation that can be use for estimation or prediction. According to Wikipedia, logistic regression analysis is a kind of regression analysis commonly use when the dependent variable is dichotomous, why we are employing logistic analysis is because the dependent variable is dichotomous i.e is having more than one value or code. An explanation of logistic regression begins with an explanation of the logistic function, which always takes on values between zero and one:

\[
\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}
\]

(2.1)

\[
\beta_0 = \log \frac{\pi}{1 - \pi}
\]

(2.2)

\[
\beta_1 = \frac{\sum X_i Y_i - \bar{X} \sum Y_i}{\sum X_i^2 - (\sum X_i)^2}
\]

(2.3)

\[\pi(x)\] is the probability of being a case,
\[\beta_0\] is the intercept from the linear regression equation (the value of the criterion when the predictor is equal to zero),
\[\beta_1 x\] is the regression coefficient multiplied by some values of the predictors,
Case e denotes the exponential function.

A. ASSUMPTIONS OF LOGISTIC REGRESSION

i. Logistic regression does not assume a linear relationship between the dependent and independent variables.

ii. The dependent variable must be a dichotomy (2 categories).

iii. The independent variables need not be interval, nor normally distributed, nor linearly related, nor of equal variance within each group.

iv. The categories (groups) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups.

v. Larger samples are needed than for linear regression because maximum likelihood coefficient are large sample estimates. A minimum of 50 cases per predictor is recommended.

B. THE LIKELIHOOD RATIO TEST

This tests the difference between \(-2\times\text{LogLikelihood}(-2LL)\) for the full model with predictors and \(-2\times\text{LogLikelihood}\) for initial chi-square in the null model. When probability fails to reach the 5% significance level, we retain the null hypothesis that knowing the independent variables (predictors) has no increased effects (i.e. make no difference) in predicting the dependent.

III. RESULTS ANALYSIS

A. DESCRIPTION AND DEMOGRAPHIC OF THE DATA

Table 1 shows that out of 200 respondents, 57.5% are male and 42.5% are female. Similarly, Table 2 reveals that the respondents are categorized into different age groups, ranging between 15 and 35 years. This will help in determining the age group where the scourge is more paramount. Contacting disease at time may due to lack of knowledge or due to ignorance on the part of the respondent. Thus, the educational status of respondent are look into. Table 3 reveals that 5.5% are having primary certificate, 25.0% are having secondary certificate, 42.5% tertiary and those that did not go to at all, are 27.0%. Also table 4 tells us that out of 200 respondents, 24.5% are living in density area, 35.5% are living a less density area, and 40.0% are living in free area. However, table 4 shows that 68.0% are having more than one sexual partner and 32.0% does not have more one sexual partner (which indicate that does that have more than one partner are more at higher percentage).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>Male</td>
<td>115</td>
<td>57.5</td>
<td>57.5</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>85</td>
<td>42.5</td>
<td>42.5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>200</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age group</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>15 – 25</td>
<td>67</td>
<td>33.5</td>
<td>33.5</td>
</tr>
<tr>
<td></td>
<td>26 – 35</td>
<td>103</td>
<td>51.5</td>
<td>85.0</td>
</tr>
<tr>
<td></td>
<td>36 – 45</td>
<td>30</td>
<td>15.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>200</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 3 Distribution of Respondents by Educational Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Status</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>Valid Primary</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
<tr>
<td>Tertiary</td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
Figure 1 shows that the scourge of Syphilis is more paramount between the age group of 26 and 35 years, with the females recording the highest number of cases, slightly bit higher than the males.

B. Omnibus Tests of Model Coefficient

Table 5 Omnibus Tests of Model Coefficient

<table>
<thead>
<tr>
<th>Step</th>
<th>Chi-square</th>
<th>Df</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>225.290</td>
<td>3</td>
<td>.000</td>
</tr>
<tr>
<td>Block</td>
<td>225.290</td>
<td>3</td>
<td>.000</td>
</tr>
<tr>
<td>Model</td>
<td>225.290</td>
<td>3</td>
<td>.000</td>
</tr>
</tbody>
</table>

-2LL table shows that out of 3 iterations the least likelihood value is 50.346 and about 90.4% of total variation in log of odd or dependent variable is explained by the predictors.

C. Hosmer and Lemeshow Test

Table 6, Hosmer and Lemeshow Test

<table>
<thead>
<tr>
<th>Step</th>
<th>Chi-square</th>
<th>DF</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.855</td>
<td>6</td>
<td>0.2</td>
</tr>
</tbody>
</table>
H0: there is a linear relationship between the dependent variable and independent variable
H1: there is no relationship between the dependent variable and independent variable.

The chi-square value is greater than \( p > 0.05 \) so we accept the null hypothesis. This is a test of the goodness of fit, similar to a chi-square test; it tests for a linear relationship between the predictor variables and log odd of the dependent variable. The table above shows a chi-square value of (7.855) with 6 degree of freedom which greater the \( p \) value so we accept the null hypothesis and conclude that there is no significant difference between the log odd of the dependent variable and independent variable.

Table 7. Variables in the Equation

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Df</th>
<th>Sig.</th>
<th>EXP(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QM</td>
<td>1.762</td>
<td>0.0116</td>
<td>1</td>
<td>.000</td>
<td>5.12</td>
</tr>
<tr>
<td>QL</td>
<td>-1.787</td>
<td>0.030</td>
<td>1</td>
<td>.114</td>
<td>.168</td>
</tr>
<tr>
<td>QN</td>
<td>-0.306</td>
<td>0.0837</td>
<td>1</td>
<td>.715</td>
<td>.757</td>
</tr>
<tr>
<td>constant</td>
<td>-8.036</td>
<td>0.0940</td>
<td>1</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 4 shows that out of the variables included in the model, QM is the only variable that is significant and this specified patient with sore throat, rash on palm of hand and swollen lymph hand, the odd ratio of the variable is 5.12 meaning that the chances or odds are 5.12 times higher to contact syphilis than ordinary person without sore throat, rash on the palm of hand, swollen lymph hand.

IV. CONCLUSION

This work examined the prevalence, demographic and estimated factors associated with people living with syphilis in Birnin Kebbi. Data were collected through questionnaire and administered at random in four different hospitals in Birnin Kebbi. The four hospitals are Sir Yahaya Memorial Hospital, Godiya Memorial Hospital, Federal Medical Centre, and Fati Lami Memorial Hospital. The method used in this research work is logistic regression methodology. The result of the descriptive statistics showed that the scourge of Syphilis is much more paramount between the age group of 26 and 35 years, with the females recording the highest number of cases, slightly bit higher than the males. The result of the logistic regression revealed that the odd ratio of the respondents with symptoms such as sore throat, swollen lymph hand and rash on the palm of their hands is 5.12, meaning that the chances or odds of contacting syphilis are 5.12 times higher for these respondent than ordinary person without these symptoms.

REFERENCES


