Log-Linear Modelling of Effect of Age and Gender on the Spread of Hepatitis B Virus Infection in Lagos State, Nigeria

Oluwole A. Odetunmibi¹, Adebowale O. Adejumo², Timothy A. Anake¹

¹Department of Mathematics, College of Science and Technology, Covenant University, Ota, Nigeria; ²Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, Kwara State, Nigeria

Abstract

BACKGROUND: The effect of age and gender on the transmission of any infectious disease can be of great importance because the age at which the host contact the disease may be a determinant on the rate at which the disease will spread.

AIM: The purpose of this research is to model the significant effect of age and gender on the spread of hepatitis B virus using data collected from Lagos State, Nigeria.

MATERIAL AND METHODS: The data that was used for this research is a ten years data covering the period of 2006 to 2015, which was collected from Nigeria Institute of Medical Research (NIMR). A log-linear modelling approach was employed using R programming language software. Akaike Information Criterion (AIC) method of model selection was used in selecting the best model.

RESULTS: It was discovered from the analysis that both factors (age and gender) have a significant effect on the spread of hepatitis B infection. This means that the age at which an individual is tested positive to hepatitis B virus will affect the spread of the disease. In choosing the best model among the four models that were developed, model AY: GY (age & year: gender and year) was found to be the best model.

CONCLUSION: Age and gender were found to act as a risk influencer that could have a great effect on the transmission of hepatitis B virus infections in Lagos state, Nigeria.

Introduction

The word hepatitis comes from the Ancient Greek word ἥπαρ (root word heat) meaning ‘liver’, and the Latin ills meaning inflammation. Hepatitis, therefore, means injury to the liver with inflammation of the liver cells [1], [2].

According to the World Health Organization, “Hepatitis is an inflammation of the liver. The condition can be self-limiting or can progress to liver cancer. Hepatitis viruses are the most common cause of hepatitis in the world, but other infections, toxic substances (e.g. alcohol, certain drugs), and autoimmune diseases may also cause hepatitis”.

It has been discovered from the literature that there are various types of Hepatitis [3] classified these various types into five major categories, namely: A, B, C, D and E, respectively. These viruses are not related to each other. They differ in their structure, the way they spread among individuals, the severity of symptoms they can cause, the way they are treated, and the outcome of the infection [3]. Among these categories, the most dangerous is Hepatitis B because it leads to chronic disease condition in hundreds of millions of people [4], [5], [6].

Hepatitis B has been described as one of the major infectious diseases in the world today because over 750,000 deaths are attributed to it annually [7], [8], [9]. Hepatitis B infection is a global healthcare problem with particularly high prevalence in developing countries in sub-Saharan Africa and South-East/Central Asia. Statistics shows that, about
350-400 million individuals worldwide suffer from chronic Hepatitis B virus infection which is a dominant cause of cirrhosis and hepatocellular carcinoma (HCC): [10], [11], [12], [13], [14], [15], [16], [17].

It has been discovered, that age and gender may be risk factors in the transmission of hepatitis B virus [18], [19], [20]. This study aimed to investigate the influence of age and gender on the prevalence of hepatitis B infections among the people of Lagos State, Nigeria. By clearly indicating the characteristics of hepatitis B and its associated risk factors, we intend to develop log-linear models and choose the best model among the developed models.

Material and Methods

The log-linear model which is used in the analysis of contingency tables is a generalised linear model for counted data, and the variety of associations and interaction terms in log-linear models can easily be described by the goodness of fit tests. The methodology of the log-linear model for the analysis of contingency tables is described in many articles and book such as [21], [22], [23], [24], [25].

Consider an I x J contingency table. The log-linear model is represented by:

\[ \log(M_{ij}) = \lambda_0 + \lambda_i + \lambda_j + \lambda_{ij} \]  

For all i and j, under the constraints of the \( \lambda \) term to sum to zero over any subscript such as:

\[ \sum_{i=1}^{I} \lambda_i = 0, \quad \sum_{j=1}^{J} \lambda_j = 0, \quad \sum_{i,j} \lambda_{ij} = 0 \]  

(2)

The log-linear model given above is called the saturated model or full model for the statistical dependency between Y1 and Y2.

By analogy with analysis of variance models, we define the overall mean by:

\[ \lambda_0 = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \log M_{ij} \]  

(3)

The main effects of Y1 and Y2 by

\[ \lambda_i = \frac{1}{J} \sum_{j=1}^{J} \log M_{ij} - \lambda_0 \]  

\[ \lambda_j = \frac{1}{I} \sum_{i=1}^{I} \log M_{ij} - \lambda_0 \]  

(4)

And the two-factor effect between Y1 and Y2 by

\[ \lambda_{ij} = \log M_{ij} = (\lambda_i + \lambda_j) - \lambda_0 \]  

(5)

(6)

Then the main and two-factor effects are determined by the odds and odds ratio, and can be written by:

\[ \lambda_i = \frac{1}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} \log \frac{M_{ij}}{M_{ij}} \]  

(7)

\[ \lambda_j = \frac{1}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} \log \frac{M_{ij}}{M_{ij}} \]  

(8)

(9)

And

So that the log-linear model for the independence model is expressed by:

\[ \log M_{ij} = \lambda_0 + \lambda_i + \lambda_j \]  

(10)

\[ \log M_{ij} = \lambda_0 + \lambda_1 + \lambda_2 \]  

(11)

For an I x J x K contingency table, the saturated log-linear model for the contingency table is:

\[ \log M_{ijk} = \lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13} + \lambda_{23} + \lambda_{123} \]  

(12)

for all i, j and k.

The \( \lambda \) terms satisfy the constraints:

\[ \sum_{i=1}^{I} \lambda_i = 0, \quad \sum_{j=1}^{J} \lambda_j = 0, \quad \sum_{i,j,k} \lambda_{ijk} = 0 \]  

(13)

\[ \sum_{i=1}^{I} \lambda_{1i} = 0, \quad \sum_{j=1}^{J} \lambda_{2j} = 0, \quad \sum_{i,j,k} \lambda_{12i} = 0 \]  

(14)

\[ \sum_{i=1}^{I} \lambda_{12i} = 0, \quad \sum_{j=1}^{J} \lambda_{23j} = 0, \quad \sum_{i,j,k} \lambda_{123i} = 0 \]  

(15)

We define the \( \lambda \) terms as follows:

The overall mean is given by:

\[ \lambda_0 = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{j=1}^{J} \log M_{ijk} \]  

(16)

The main effects of Y1, Y2, and Y3 are:

\[ \lambda_i = \frac{1}{IK} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{i=1}^{I} \log M_{ijk} - \lambda_0 \]  

(17)

\[ \lambda_j = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{j=1}^{J} \log M_{ijk} - \lambda_0 \]  

(18)
Results

The summary of the data that was used for this research is presented in Table 1. The data covers the period of ten years (2006 – 2015) of those that are tested positive to hepatitis B virus in Lagos state, Nigeria.

Table 1: Classification of the data according to the attributes of age and gender of the patients and the year they were diagnosed with having the disease

<table>
<thead>
<tr>
<th>Age Interval</th>
<th>Gender</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;20 years</td>
<td>Male</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>54</td>
</tr>
<tr>
<td>21-30 years</td>
<td>Male</td>
<td>45</td>
<td>60</td>
<td>132</td>
<td>101</td>
<td>72</td>
<td>68</td>
<td>52</td>
<td>65</td>
<td>29</td>
<td>20</td>
<td>644</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>93</td>
<td>112</td>
<td>191</td>
<td>163</td>
<td>118</td>
<td>103</td>
<td>90</td>
<td>110</td>
<td>50</td>
<td>34</td>
<td>1064</td>
</tr>
<tr>
<td>31-40 years</td>
<td>Male</td>
<td>208</td>
<td>171</td>
<td>330</td>
<td>289</td>
<td>284</td>
<td>269</td>
<td>107</td>
<td>231</td>
<td>147</td>
<td>108</td>
<td>2826</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>435</td>
<td>451</td>
<td>719</td>
<td>633</td>
<td>697</td>
<td>694</td>
<td>345</td>
<td>575</td>
<td>341</td>
<td>265</td>
<td>4545</td>
</tr>
<tr>
<td>41-50 years</td>
<td>Male</td>
<td>181</td>
<td>158</td>
<td>280</td>
<td>236</td>
<td>153</td>
<td>168</td>
<td>139</td>
<td>207</td>
<td>99</td>
<td>71</td>
<td>1692</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>377</td>
<td>331</td>
<td>562</td>
<td>525</td>
<td>368</td>
<td>386</td>
<td>310</td>
<td>398</td>
<td>110</td>
<td>94</td>
<td>3461</td>
</tr>
<tr>
<td>51-60 years</td>
<td>Male</td>
<td>30</td>
<td>38</td>
<td>80</td>
<td>71</td>
<td>51</td>
<td>47</td>
<td>39</td>
<td>44</td>
<td>11</td>
<td>7</td>
<td>418</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>129</td>
<td>173</td>
<td>275</td>
<td>224</td>
<td>195</td>
<td>207</td>
<td>139</td>
<td>153</td>
<td>57</td>
<td>31</td>
<td>1583</td>
</tr>
<tr>
<td>61-70 years</td>
<td>Male</td>
<td>8</td>
<td>13</td>
<td>19</td>
<td>15</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>5</td>
<td>4</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>31</td>
<td>39</td>
<td>68</td>
<td>50</td>
<td>39</td>
<td>45</td>
<td>28</td>
<td>35</td>
<td>20</td>
<td>15</td>
<td>370</td>
</tr>
<tr>
<td>&gt;70 years</td>
<td>Male</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>35</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1552</td>
<td>1568</td>
<td>2753</td>
<td>2329</td>
<td>2012</td>
<td>2014</td>
<td>1277</td>
<td>1844</td>
<td>882</td>
<td>656</td>
<td>16887</td>
</tr>
</tbody>
</table>

Log-Linear Fitted Models

The summary of the log-linear fitted models for the data presented in table 1 concerning the effect of age, gender and the year of being tested positive is presented below:

Table 2: Model 1: Age: Gender: Year (AG: Y)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$X^2$</th>
<th>df</th>
<th>$P(&gt; X^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>235.6358</td>
<td>60</td>
<td>0.0034421</td>
</tr>
<tr>
<td>Pearson</td>
<td>299.4350</td>
<td>60</td>
<td>0.0110223</td>
</tr>
</tbody>
</table>

Model 1 presented in table 2 above considered when no association exists among the variables under consideration. The model is written mathematically as:

$$\log(m_{ijk}) = \lambda + \lambda^A + \lambda^G + \lambda^Y$$

(24)

Table 3: Model 2: Age and Year: Gender and Year (AY: GY)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$X^2$</th>
<th>df</th>
<th>$P(&gt; X^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>227.5998</td>
<td>69</td>
<td>0.081745</td>
</tr>
<tr>
<td>Pearson</td>
<td>224.8027</td>
<td>69</td>
<td>0.034568</td>
</tr>
</tbody>
</table>

Model 2 presented in table 3 above considered two-way association between each variable with the year of being tested positive. The model is written mathematically as:

$$\log(m_{ijk}) = \lambda + \lambda^A + \lambda^G + \lambda^Y + \lambda^AY + \lambda^GY$$

(25)

Table 4: Model 3: Age and Gender: Year (AG: Y)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$X^2$</th>
<th>df</th>
<th>$P(&gt; X^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>49.95058</td>
<td>54</td>
<td>0.031274</td>
</tr>
<tr>
<td>Pearson</td>
<td>49.97675</td>
<td>54</td>
<td>0.030274</td>
</tr>
</tbody>
</table>

Model 3 presented in table 4 above considered two-way association between the two variables together the year of being tested positive. The model is written mathematically as:

$$\log(m_{ijk}) = \lambda + \lambda^A + \lambda^G + \lambda^Y + \lambda^AG + \lambda^GY$$

(26)

Table 5: Model 4: Age: Gender: Year (A: G: Y)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>$X^2$</th>
<th>df</th>
<th>$P(&gt; X^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>49.95058</td>
<td>54</td>
<td>0.031274</td>
</tr>
<tr>
<td>Pearson</td>
<td>49.97675</td>
<td>54</td>
<td>0.030274</td>
</tr>
</tbody>
</table>

Model 4 presented in table 4 above considered three-way association among the variables. The model is written mathematically as:

$$\log(m_{ijk}) = \lambda + \lambda^A + \lambda^G + \lambda^Y + \lambda^AG + \lambda^GY + \lambda^AY + \lambda^GY$$

(27)

Table 6: AIC Values for The Models

<table>
<thead>
<tr>
<th>MODEL</th>
<th>$G^2$</th>
<th>AIC</th>
<th>$P$-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>72.2814</td>
<td>122.05</td>
<td>0.1981745</td>
</tr>
<tr>
<td>Model 2</td>
<td>236.6358</td>
<td>117.37</td>
<td>0.0034421</td>
</tr>
<tr>
<td>Model 3</td>
<td>227.5998</td>
<td>122.67</td>
<td>0.081745</td>
</tr>
<tr>
<td>Model 4</td>
<td>49.95058</td>
<td>127.04</td>
<td>0.031274</td>
</tr>
</tbody>
</table>

Discussion

Comparing the $P$-values of all the models with 0.05 level of significance; model 1, which is the model that represents no association among the variables under consideration, is the only model that is not significant. Model 2, 3 and 4 that established an interaction among the variables are all significant. Therefore, modelling the effect of age and gender on the spread of HBV virus infection in Lagos state for the period of ten years of 2006 to 2015 shows that, both variables (age and gender) have a significant effect on the spread of the disease (Table 2, 3 and 4). On the other hand, choosing the best model among various log-linear models developed model:
age and year: gender and years was discovered to be the best model since the AIC value (117.37) for the model are the lowest and the highest using likelihood ratio test (235.63583) (Table 6). This means that age and gender are not independent of the spread of the disease that is, the effect of age on the spread of the disease is not independent of the effect of gender on the spread of the disease.

Acknowledgement

The authors wish to appreciate the efforts and commitment of the anonymous reviewers toward this publication. The financial support from Covenant University, Nigeria, is also deeply appreciated.

References