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

Oluwole A. Odetunmibi[1*], Adebowale O. Adejumo[2], Timothy A. Anake[1]

[1]*Department of Mathematics, College of Science and Technology, Covenant University, Ota, Nigeria;* [2]*Department 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 important 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 *hepar* (root word *heat*) meaning 'liver', and the Latin *itis* 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:

$$\log(M_{ij}) = \lambda_0 + \lambda_i^1 + \lambda_j^2 + \lambda_{ij}^{12} \quad (1)$$

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

$$\sum_{i=1}^{I} \lambda_i^1 = 0, \qquad \sum_{j=1}^{J} \lambda_j^2 = 0,$$
$$\sum_{i=1}^{I} \lambda_{ij}^{12} = \sum_{j=1}^{J} \lambda_{ij}^{12} = 0 \quad (2)$$

The log-linear model given above is called the saturated model or full model for the statistical dependency between Y_1 and $Y2$.

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

$$\lambda_0 = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \log M_{ij} \quad (3)$$

The main effects of Y_1 and Y_2 by

$$\lambda_i^1 = \frac{1}{J} \sum_{j=1}^{J} \log M_{ij} - \lambda_0 \quad (4)$$

$$\lambda_i^1 = \frac{1}{I} \sum_{j=1}^{J} \log M_{ij} - \lambda_0 \quad (5)$$

And the two-factor effect between Y_1 and Y_2 by

$$\lambda_{ij}^{12} = \log M_{ij} = (\lambda_i^1 + \lambda_j^2) - \lambda_0 \quad (6)$$

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

$$\lambda_i^1 = \frac{1}{IJ} \sum_{i^1=1}^{I} \sum_{j=1}^{J} \log \frac{M_{ij}}{M_{i^1 j}}, \quad (7)$$

$$\lambda_i^2 = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j^1=1}^{J} \log \frac{M_{ij}}{M_{ij^1}} \quad (8)$$

And

$$\lambda_{ij}^{12} = \frac{1}{IJ} \sum_{i^1=1}^{I} \sum_{j^1=1}^{J} \log \frac{M_{ij} M_{ij}}{M_{i^1 j} M_{ij^1}} \quad (9)$$

For the independence model that Y_1 is statistically independent of Y_2, the cell probability Mij can be factorised into the product of marginal probabilities M_{i+} and M_{+j}, that is,

Mij = Mi+ M+j,

Where $M_{i+} = \sum_{j=1}^{J} M_{ij}$ and $M_{+j} = \sum_{i=1}^{I} M_{ij}$.

Then the two-factor effect is:

$$\lambda_{ij}^{12} = \frac{1}{IJ} \sum_{i^1=1}^{I} \sum_{j^1=1}^{J} \log \frac{M_{i+} M_{+j} M_{i^1+} M_{+j^1}}{M_{i^1+} M_{+j} M_{i+} M_{+j^1}} = 0 \quad , (10)$$

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

$$\log M_{ij} = \lambda_0 + \lambda_i^1 + \lambda_j^2 \text{, for all I and j} \quad (11)$$

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

$$\log M_{ijk} = \lambda_0 + \lambda_i^1 + \lambda_j^2 + \lambda_k^3 + \lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23} + \lambda_{ijk}^{123} \quad (12)$$

for all i, j and k.

The λ terms satisfy the constraints;

$$\sum_{i=1}^{I} \lambda_i^1 = \sum_{j=1}^{J} \lambda_j^2 = \sum_{k=1}^{K} \lambda_k^3 = 0, \quad (13)$$

$$\sum_{i=1}^{I} \lambda_{ij}^{12} = \sum_{j=1}^{J} \lambda_{ij}^{12} = ... = \sum_{k=1}^{K} \lambda_{jk}^{23} = 0, \quad (14)$$

$$\sum_{i=1}^{I} \lambda_{ijk}^{123} = \sum_{j=1}^{J} \lambda_{ijk}^{123} = \sum_{k=1}^{K} \lambda_{ijk}^{123} = 0, \quad (15)$$

We define the λ terms as follows:

The overall mean is given by:

$$\lambda_0 = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \log M_{ijk.} \quad (16)$$

The main effects of Y_1, Y_2, and Y_3 are:

$$\lambda_i^1 = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \log M_{ijk.} - \lambda_0 \quad (17)$$

$$\lambda_j^2 = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} \log M_{ijk.} - \lambda_0 \quad (18)$$

$$\lambda_k^2 = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \log M_{ijk.} - \lambda_0 \quad (19)$$

Each interaction effect is given by:

_____

$$\lambda_{ij}^{12} = \frac{1}{K} \sum_{k=1}^{K} \log M_{ijk.} - (\lambda_i^1 + \lambda_j^2) - \lambda_0 \quad (20)$$

$$\lambda_{ik}^{13} = \frac{1}{J} \sum_{j=1}^{J} \log M_{ijk.} - (\lambda_i^1 + \lambda_k^3) - \lambda_0 \quad (21)$$

$$\lambda_{jk}^{23} = \frac{1}{I} \sum_{i=1}^{I} \log M_{ijk.} - (\lambda_j^2 + \lambda_k^3) - \lambda_0 \quad (22)$$

and,

$$\lambda_{ijk}^{123} = \log M_{ijk.} - (\lambda_{ij}^{12} + \lambda_{ik}^{13} + \lambda_{jk}^{23}) - (\lambda_i^1 + \lambda_j^2 + \lambda_k^3) - \lambda_0 \quad (23)$$

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

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

Log-Linear Fitted Models

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

Table 2: Model 1: Age: Gender: Year (A: G: Y)

Statistics:	X^2	df	P (> X^2)
Likelihood Ratio	72.28144	63	0.1981745
Pearson	73.03648	63	0.1815950

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

$$\log(m_{ijk}) = \lambda + \lambda_i^A + \lambda_j^G + \lambda_k^Y \quad (24)$$

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

Statistics:	X^2	df	P (> X^2)
Likelihood Ratio	235.6358	60	0.0034421
Pearson	299.4320	60	0.0110223

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:

$$\log(m_{ijkl}) = \lambda + \lambda_i^A + \lambda_j^G + \lambda_k^Y + \lambda_{ik}^{AY} + \lambda_{jk}^{GY} \quad (25)$$

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

Statistics:	X^2	df	P (> X^2)
Likelihood Ratio	227.5998	69	0.081745
Pearson	224.8027	69	0.034568

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:

$$\log(m_{ijkl}) = \lambda + \lambda_i^A + \lambda_j^G + \lambda_k^Y + \lambda_{ik}^{AG} \quad (26)$$

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

Statistics:	X^2	df	P (> X^2)
Likelihood Ratio	49.95058	54	0.0312744
Pearson	49.97675	54	0.0302744

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

$$\log(m_{ijkl}) = \lambda + \lambda_i^A + \lambda_j^G + \lambda_k^Y + \lambda_{ij}^{AG} + \lambda_{ik}^{AY} + \lambda_{jk}^{GY} \quad (27)$$

Table 6: Aic Values for The Models

MODEL	G²	AIC	P-VALUE
Model 1	72.28144	122.05	0.1981745
Model 2	235.63583	117.37	0.0034421
Model 3	227.59983	122.67	0.081745
Model 4	49.95058	127.04	0.0312744

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, 4 and 5).

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

References

1. Karthikeyan T. Analysis of classification algorithms applied to hepatitis patients. International Journal of Computer Applications. 2013; 62(15):2530. https://doi.org/10.5120/10157-5032

2. Naghavi M, Wang H, Lozano R, et al. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 385:117-71. https://doi.org/10.1016/S0140-6736(14)61682-2

3. World Health Organization. Hepatitis B fact Sheet No. 204, 2014. http://www.who.int/mediacentre/factsheets/fs204/en/

4. Ikobah J, Okpara H, Elemi I, Ogarepe Y, Udoh E, Ekanem E. The prevalence of hepatitis B virus infection in Nigerian children prior to vaccine introduction into the National Programme on Immunization schedule. Pan African Medical Journal. 2016; 23(1):128. https://doi.org/10.11604/pamj.2016.23.128.8756 PMid:27279955 PMCid:PMC4885699

5. White OD, Fenner JF. Viruses of humans, medical virology (4th ed.) Academic Press Ltd., 1994.

6. Onuzulike N, Ogueri EO. Sero-prevalence of hepatitis B surface antigen (HBsAg) in pregnant women in Owerri, Imo State of Nigeria. Res J Biol Sci. 2007; 2:178-82. https://doi.org/10.4314/ijonas.v3i2.36178

7. Maraolo AE, Gentile I, Buonomo AR, Pinchera B, Borgia G. Current evidence on the management of hepatitis B in pregnancy. World Journal of Hepatology. 2018; 10(9):585-594. https://doi.org/10.4254/wjh.v10.i9.585 PMid:30310536 PMCid:PMC6177570

8. World Health Organization. (2009). Hepatitis B fact Sheet No. 204. http://www.who.int/mediacentre/factsheets/fs204/en/

9. WHO. Draft global health sector strategies. Viral hepatitis, 2016-2021. Report by the Secretariat. 2016. http://apps.who.int/gb/ebwha/pdf_files/WHA69/A69_32-en.pdf?ua=1 (accessed 10 Oct 2016).

10. Lavanchy, D. Worldwide epidemiology of HBV infection, disease burden, and vaccine prevention. Journal of Clinical Virology. 2005; 34(1):S1-S3. https://doi.org/10.1016/S1386-6532(05)00384-7

11. Ola SO, Odaibo GN. Alpha-fetoprotein, HCV and HBV infections in Nigerian patients with PHCC. 2007; 51(3):33-35. https://doi.org/10.4314/nmp.v51i3.28903

12. Hebo HJ, Gemeda DH, Abdusemed KA. Hepatitis B and C Viral Infection: Prevalence, Knowledge, Attitude, Practice, and Occupational Exposure among Healthcare Workers of Jimma University Medical Center, Southwest Ethiopia. The Scientific World Journal. 2019:195-206. https://doi.org/10.1155/2019/9482607 PMid:30853866 PMCid:PMC6377947

13. Ray G. Current Scenario of Hepatitis B and its Treatment in India. Journal of Clinical and Translational Hepattology. 2017; 5(3):277-296. https://doi.org/10.14218/JCTH.2017.00024 PMid:28936409 PMCid:PMC5606974

14. Zou L, Zhang W, Ruan S. Modeling the transmission dynamics and control of hepatitis B virus in China. Journal of Theor Bio. 2010; 262:330-338. https://doi.org/10.1016/j.jtbi.2009.09.035 PMid:19822154

15. Zhu J, Yang W, Feng Y, Lo C, Chen H, Zhu Q, Shen Z, Lan G, Chen Y, Tang Z, Xing H. Treatment effects of the differential first-line antiretroviral regimens among HIV/HBV coinfected patients in southwest China: an observational study. Scientific reports. 2019; 9(1):1006-1025. https://doi.org/10.1038/s41598-018-37148-8 PMid:30700732 PMCid:PMC6353888

16. Anahi EM. Prevalence and risk factors of hepatitis B virus infection in Bahrain, 2000 through 2010. PLoS One. 2014; 9:e87599. https://doi.org/10.1371/journal.pone.0087599 PMid:24498341 PMCid:PMC3911996

17. Ola SO, Odaibo GN, Olaleye OD, Ayoola EA. Hepatitis B and E viral infections among Nigerian healthcare workers. Afr J Med Medic Sci. 2013; 41(4):387-391.

18. Adejumo AO, Ikoba NA, Suleiman EA, Okagbue HI, Oguntunde PE, Odetunmibi OA, Job O. Quantitative exploration of factors influencing psychotic disorder ailments in Nigeria. Data in Brief. 2017; 14:175-85. https://doi.org/10.1016/j.dib.2017.07.046 PMid:28795095 PMCid:PMC5537424

19. Adamu PI, Oguntunde PE, Okagbue HI, Agboola OO. On the Epidemiology and Statistical Analysis of HIV/AIDS Patients in the Insurgency Affected States of Nigeria. Open Access Maced J Med Sci. 2018; 6(7):1315-1321. https://doi.org/10.3889/oamjms.2018.229 PMid:30087744 PMCid:PMC6062286

20. Yang S, Cheng D, Yuanxia C, Jie W, Chengbo Y, Ping C, Kaijin X, Min D, Yiping L, Juanjuan L, Pei Y, Wen R, Yan Q, Qing C, Yuqing Z, Jun Y, Bing R, Jingjing R and Lanjuan L. Prevalence and influencing factors of hepatitis B among a rural residential population in Zhejiang Province, China: a cross-sectional study. BMJ Open. 2017; 7:e014947. https://doi.org/10.1136/bmjopen-2016-014947 PMid:28373257 PMCid:PMC5387940

21. Oguntunde PE, Adejumo AO, Okagbue HI. Breast Cancer Patients in Nigeria: Data exploration approach. Data in Brief. 2017; 15:47-57. https://doi.org/10.1016/j.dib.2017.08.038 PMid:28971122 PMCid:PMC5612794

22. Adejumo AO. Modelling Generalized Linear (Loglinear) Models for Raters Agreement measures. Peter Lang, Frankfurt am Main: Germany, 2005.

23. Agresti A. Categorical Data Analysis. John Wiley and Sons. 2nd Edition. New York, 2002. https://doi.org/10.1002/0471249688

24. Everitt BS. The analysis of contingency tables. Chapman and Hall/CRC; 1992. https://doi.org/10.2307/2532373

25. Odetunmibi OA, Adejumo AO, Sanni OOM. Loglinear Modelling of Cancer Patients Cases in Nigeria: An Exploratory Study Approach. Open Science Journal of Statistics and Application. 2013; 1(1):1-7.