



INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY
**Binary Logistic Regression Analysis on Risk Factors Associated with
 Hepatitis B Disease in Gusau Local Government Zamfara State,
 Nigeria**

Tolulope Olayemi James¹, Gerald Ike Onwuka² and Usman Bello³

Department of Mathematics, Kebbi State University of Science and Technology, Aliero, Kebbi State, Nigeria

Corresponding Author's E-mail and Phone: usmanbellogusau@gmail.com 08036477128

Received on: August, 2023

Revised and Accepted on: September, 2023

Published on: October, 2023

ABSTRACT

This study focused on creating a binary logistic regression model to evaluate the risk factors for hepatitis B viral disease in the Gusau Local Government of Zamfara State, Nigeria. Secondary data was utilised, and the model identified independent predictors of hepatitis-b disease based on variables with a P-value < 0.05 and a 95% confidence interval for the odds ratio not less than 1. The results indicated that marital status, history of blood transfusion, multiple sex partners, and alcohol consumption were statistically associated with hepatitis-B virus infection, while gender and having an infected family member were not. Model selection criteria, including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), favoured a model with only significant predictors, revealing the existence of Hepatitis-B virus infection risk factors in patients from Gusau Local Government of Zamfara State.

Keywords: Hepatitis B, Logistic Regression, Odds ratio, Infection, Risk Factor, Diabetes

1.0 INTRODUCTION

Hepatitis B (HBV) stands as a formidable global health threat, giving rise to potentially life-threatening liver infections that carry the risk of cirrhosis and liver cancer. Effectively preventing HBV infection is paramount to sidestepping complications, including the development of chronic diseases and liver cancer. The World Health Organization's (WHO) data from 2021 underscores the substantial burden of hepatitis B, with the Western Pacific and African Regions shouldering the highest numbers, counting 116 million and 81 million people, respectively, facing chronic infections. Within this context, Nigeria emerges as a hyper-endemic hotspot, grappling with an 8.1% prevalence, impacting an estimated 19 million individuals according to the Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS, 2018).

In regions highly endemic for HBV, transmission patterns often involve perinatal transmission from mother to child or horizontal transmission during the initial five years of life. This horizontal transmission occurs through close body contact, especially with children in an active infection or carrier state. Factors such as skin lesions, like

impetigo, scabies, and cuts, facilitate the transfer of blood and bodily fluids. Moreover, sharing personal items such as toothbrushes, towels, beds, razors, needles, nail and hair clippers contributes to the spread of the virus (WHO, 2020).

Given Nigeria's substantial population of 212 million, coupled with relatively high prevalence rates of hepatitis B and hepatitis C, the country assumes a pivotal role in global efforts to eliminate hepatitis. The gravity of the situation is underscored by the fact that HIV and hepatitis, according to Global Burden of Disease estimates in 2016, were directly linked to over 1.6 million deaths and 59 million years of life lost. The prevalence of hepatitis in Nigeria, especially in Zamfara State, has manifested in critical epidemiological and clinical outcomes. Recent studies highlight that in 2021, hepatitis B alone accounted for up to 10% of hospital admissions in Nigeria (WHO, 2021). While the prevalence of hepatitis in Nigeria may not be directly comparable to diseases like malaria, the economic toll of caring for hepatitis far exceeds the costs associated with treating many other communicable diseases. This reinforces the urgency and significance of addressing the burden of hepatitis B

comprehensively in the healthcare landscape of Nigeria.

2.0 LITERATURE REVIEW

Datong et al. (2017) conducted a study exploring cardiovascular disease prevalence and associated risk factors, focusing on hepatitis B virus infection in a sample of 693 patients at Plateau State Specialist Hospital, Jos, Nigeria. Their logistic regression analysis involving age, history of blood transfusion, and alcohol consumption revealed a moderate relationship between these risk factors and Hepatitis B Virus infection. Mohd et al. (2019) investigated risk factors linked to hepatitis C infection in adult patients in Kedah state, Malaysia. Using self-administered questionnaires, they gathered data on demographic characteristics and exposure to selected risk factors. Univariate and multivariable logistic regression analyses showed independent associations between hepatitis C and factors like blood transfusion history, injection drug use, imprisonment, tattooing, and having multiple sexual partners.

Anteneh et al. (2021) employed a binary logistic regression model to identify predictors of hepatitis B virus infection in HIV-positive pregnant women in Eastern Amhara region, Ethiopia. Their study, utilizing interview and chart review techniques, highlighted statistically significant associations between hepatitis B infection and factors such as a history of sexually transmitted infections, family history of HBV, and unsafe sex.

Zhang et al. (2011) investigated risk factors for Hepatitis B infection in Northeast China through a cross-sectional survey involving 3833 people. Multivariate logistic regression analysis identified independent predictors for hepatitis infection, including large family size, occupation as a private small-businessman, male gender, family history of HBV, personal history of vaccination, and older age.

3.0 METHODOLOGY

The research analysis relies on data obtained from the Gastro clinic department of the Federal Medical Centre in Gusau, Zamfara State. The data was collected from January 2018 to December 2019 and encompasses various information such as

demographic details (marital status, age), factors like blood transfusion and sharing sharp objects, alcohol consumption, and the utilisation of patients' records.

3.1. Logistic Regression Model

The logistic regression model resembles the regression equation, especially when dealing with a single predictor variable, X_1 . In this scenario, the logistic regression equation predicts the probability of Y as follows:

$$\pi(x) = \frac{\ell^{\beta_0 + \beta_1 x}}{1 + \ell^{\beta_0 + \beta_1 x}} \quad (1)$$

In which $\pi(x)$ is the probability of Y occurring, e is the base of natural logarithms, and the other coefficients form a linear combination much the same as in simple regression. This equation can be extended to include many predictors, just like linear regression. When more than one predictor is present, the equation becomes:

$$\pi(x) = P(Y = 1 / X_1, X_2, \dots, X_K) = \frac{\ell^{\beta_0 + \sum_{j=1}^k \beta_j X_j + \varepsilon}}{1 + \ell^{\beta_0 + \sum_{j=1}^k \beta_j X_j + \varepsilon}} \quad (1)$$

The above equation (1) is the same as the equation (2) when just one predictor is present, except that the linear combination has been expanded to include any number of predictors.

3.2 Assessment of Fitted Logistic Regression Model

After determining the coefficients, various assessment parameters or tests should be conducted to scrutinize the suitability, relevance, and sufficiency of the logistic regression model produced. Examples of these evaluation parameters include statistical tests for each predictor variable, goodness-of-fit statistics, and model discrimination.

3.2.1 The Wald Statistic

The Wald Statistic analogous to the t-test in linear regression is used to assess the significant of individual logistic regression coefficients. It is the

ratio of the square of regression coefficient to the square of the standard error of the coefficient.

$$W_j = \frac{\widehat{\beta}_j^2}{SE(\widehat{\beta}_j)^2} \quad (3)$$

Where $\widehat{\beta}_j$ represents the estimated coefficient of β and $SE(\widehat{\beta}_j)$ is its standard error. Under the null hypothesis $H_0: \beta_j = 0$. The quantity follows a chi-square distribution with one degree of freedom.

3.2.2 Goodness of Fit the Hosmer-Lemeshow Test

Hosmer and Lemeshow (2000) developed a goodness-of-fit test for logistic regression models with binary responses. They proposed grouping based on the value of the estimated probabilities. This test is obtained by calculating the Pearson chi-square statistic from the $2 \times g$ table of observed and expected frequencies, where g is the number of groups. The statistic is written as:

$$H = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{E_g} \quad (4)$$

Where, O_g and E_g denote the observed events and expected events for the g th risk deciles groups, respectively. This test follows χ^2 distribution with 8 (number of groups-2) degree of freedom.

3.3 Odd Ratio

The odds ratio is a measure of association for 2×2 contingency table (Agresti, 2007). In 2×2 tables Agresti (2002) define the odds ratio in two groups of subjects as "the ratio of odds". Thus;

$$OR = \frac{odds_1}{odds_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \quad (5)$$

Hence, for logistic regression with a dichotomous independent variable coded 1 and 0, the relationship between the odds ratio and the regression coefficient is $OR = e^{\beta_i}$

A 95% CI for log Odds ratio is given by:

$$\text{Log Odds ratio} = \ln(OR) \pm 1.96 \times \{SE \ln(OR)\} \quad (6)$$

Where, $\ln(OR)$ is the sample log odds ratio and $SE \ln(OR)$ is the standard error of the log odds ratio.

A 95% Confidence interval for Odds ratio is given by:

$$\text{Odds ratio} = e^{\ln(OR) \pm 1.96 \times [SE \ln(OR)]} \quad (7)$$

3.4 Model Selection Criterion

3.4.1 Akaike Information Criterion (AIC)

Assuming we have a set of statistical models for the data, the favored model is the one with the lowest AIC value. AIC prioritizes the goodness of fit, evaluated through the likelihood function. The AIC value for the model is as follows.

$$AIC = 2k - 2 \ln(\bar{L})$$

Where, k = number of estimated parameters in the model.

\bar{L} = Maximized likelihood function for the estimated model

3.4.2 Bayesian Information Criterion (BIC)

In statistics, the Bayesian information criterion (BIC) serves as a measure for selecting the most suitable model from a finite array. Models exhibiting lower BIC values are typically favored. This criterion relies partially on the likelihood function and shares close ties with the Akaike information criterion (AIC). Gideon E. Schwarz introduced the BIC in 1978, providing a Bayesian rationale for its adoption. The formal definition of BIC is

formally defined as

$$BIC = k \ln(n) - 2 \ln(\bar{L})$$

Where

\bar{L} = The maximized value of the likelihood function of the model

n = The number of observations, or equivalently, the sample size

k = The number of parameters estimated by the model.

4.0 RESULTS AND DISCUSSION

4.1 Binary Logistic Regression Results

Examining the Wald Test and the significance column in the table indicates that the noteworthy predictors include Marital Status (MS) with a p-

value of 0.021 (< 0.05), History of Blood Transfusion (HBT) with a p-value of 0.000 (< 0.05), Multiple Sexual Partner (MSP) with a p-value of 0.006 (< 0.05), and Alcohol consumption (AC) with a p-value of 0.001 (< 0.05). Conversely, Gender (p-value = 0.123 > 0.05) and Infected Family Member (IFM) (p-value = 0.196 > 0.05) are predictors that lack statistical significance.

The Fitted Binary Logistic Regression Model with All Parameters

$$\ln\left(\frac{\theta}{1-\theta}\right) = -1.156 + 0.661 * MS + 1.073 * HBT + 0.811 * MSP + 0.955 * AC - 0.454 * Gender - 0.374 * IFM + \varepsilon$$

Table 1: Model 1 - Analysis of the Maximum Likelihood Estimate of model with all Predictors

| Parameter | Estimate | Standard Error | Wald Test | Sig. |
|------------------------------------|----------|----------------|-----------|-------|
| Intercept | -1.156 | 0.307415 | 14.129 | 0.000 |
| Marital Status(MS) | 0.661 | 0.287414 | 5.288 | 0.021 |
| History of Blood Transfusion (HBT) | 1.073 | 0.290956 | 13.605 | 0.000 |
| Multiple Sexual Partner(MSP) | 0.811 | 0.292831 | 7.675 | 0.006 |
| Alcohol consumption(AC) | 0.955 | 0.287038 | 11.067 | 0.001 |
| Gender | -0.454 | 0.294680 | 2.378 | 0.123 |
| Infected Family Member (IFM) | -0.374 | 0.289256 | 1.669 | 0.196 |

Table 2: Model 2-Analysis of the maximum likelihood estimate of only significant coefficients and their standard error

| Parameter | Estimate | Standard Error | Wald Test | Sig. |
|------------------------------------|----------|----------------|-----------|-------|
| Intercept | -1.439 | 0.279 | 26.717 | 0.000 |
| Marital Status (MS) | 0.688 | 0.283 | 5.934 | 0.015 |
| History of Blood Transfusion (HBT) | 0.992 | 0.284 | 12.189 | 0.000 |
| Multiple Sexual Partner (MSP) | 0.698 | 0.284 | 6.061 | 0.014 |
| Alcohol consumption (AC) | 0.935 | 0.284 | 10.859 | 0.001 |

The Fitted Binary Logistic Regression Model with Only Significant Parameters

$$\ln\left(\frac{\theta}{1-\theta}\right) = -1.439 + 0.688 * MS + 0.992 * HBT + 0.698 * MSP + 0.935 * AC + \varepsilon$$

4.2 The Odd Ratio

Table 3 shows the exponential of parameter estimates in the Exp(B) column. The odds ratios for the respective predictor factors are shown by these variables. From Table 3, it is evident that individuals with History of Blood Transfusion and those that consume alcohol are highly susceptible for hepatitis B occurrence.

Table 3: Odds Ratios and Confidence Intervals for the coefficient estimates of the Significant Predictors

| Parameter | $Exp(\beta)$ | 95% confidence Interval for $Exp(\beta)$ | |
|------------------------------------|--------------|--|-------|
| | | Upper | Lower |
| Marital Status(MS) | 1.99 | 1.144 | 3.46 |
| History of Blood Transfusion (HBT) | 2.70 | 1.54 | 4.70 |
| Multiple Sexual Partners (MSP) | 2.01 | 1.15 | 3.50 |
| Alcohol consumption (AC) | 2.55 | 1.46 | 4.44 |

4.3 Model Assessment and Selection

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) of the two models are shown in table 4. The computed AIC and BIC of each model should be taken into account while choosing the optimum model for hepatitis B virus. The better the model fits the data, the smaller these values are. Hence, the second model, which has only significant coefficients, will be used to make predictions.

Table 4: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

| Model | AIC | BIC |
|--|---------|---------|
| Model-1 with all coefficients | 146.863 | 171.513 |
| Model-2 with only significant coefficients | 69.644 | 87.252 |

4.4 Goodness of Fit Statistics: The Hosmer-Lemeshow Test

Table 5 presents the Hosmer and Lemeshow test results, indicating a significance of 0.533. This non-significant result suggests that we do not have statistical evidence to reject the null hypothesis. Consequently, we can conclude that there is no significant difference between observed and model-predicted values, implying that the model's estimates align well with the data.

Table 5: Hosmer and Lemeshow Test

| Model | Chi-Square | Df | Sig. |
|--|------------|----|-------|
| Model-2 with only significant predictors | 7.233 | 5 | 0.204 |

Table 6: Goodness of fit

| Model | Chi-Square | Df | Sig. |
|--|------------|----|-------|
| Model-2 with only significant predictors | 12.7335 | 11 | 0.311 |

Table 6 presents the deviance analysis, indicating a significance of 0.311. This result suggests non-statistical significance, leading to the retention of the null hypothesis. Consequently, it implies no distinction between observed and model-predicted values, affirming that the model's estimates suit the data adequately.

5.0 CONCLUSION

The risk factors for getting hepatitis B infection were investigated using a logistic regression model in this study. Marital Status, Blood Transfusion History, Multiple Sexual Partners, Alcohol Consumption, Gender, and Infected Family Member are the risk factors utilized. The chance of having hepatitis B is estimated using a binary logistic regression model. To begin, the chi-square test of association between hepatitis B and all predictor variables revealed that Marital Status, Blood Transfusion History, Multiple Sexual Partners, and Alcohol Consumption are all statistically significant.

Secondly, using the Wald test to examine the significance of the logistic coefficients, it was discovered that characteristics such as marital status, blood transfusion history, multiple sexual partners, and alcohol use are significant predictors of hepatitis B. The fitted model revealed that developing hepatitis B is unaffected by a person's gender or having a history of infected family members. The maximum likelihood test, as well as the Hosmer and Lemeshow tests, were employed to evaluate the model's fitness. Finally, a model with just significant predictors was chosen for prediction because it was shown to fit the data well enough.

6.0 RECOMMENDATIONS

The following recommendations were formed based on the findings of this study in order to contribute to the fight against the most debilitating disease in individuals, hepatitis B:

The secondary data used in the study came from the Federal Medical Center in Gusau, Zamfara State. Because of the uncertainties involved with

secondary data, we urge that researchers who want to conduct study locate primary data from patients.

More follow-up studies should be conducted to assess the benefits of various treatment modalities on the control of hepatitis B risk factors such as marital status, history of blood transfusion, multiple sexual partners, alcohol consumption, gender, and infected family members in Gusau local government and, if possible, Nigeria, to prevent further complications and new infections.

The model's application should be combined with other recommendations for lowering the risk of hepatitis B infection.

Areas for further Research

The following areas were recommended for further research:

Other statistical approach such as machine bearing approach and modified poisson should be applied to this study to compare the result obtained with multivariate approach

More risk factors should be identified and studied

The scope of the study should be widen i.e to accommodate more health institutions including both public and private health institutions.

REFERENCES

- Agresti, A. (2002). *Categorical data analysis*. John Wiley & Sons, New York, 2nd edition
- Anteneh, Z. A., Wondaye, E., & Mengesha, E. W. (2021). Hepatitis B virus infection and its determinants among HIV positive pregnant women: Multicenter unmatched case-control study. *PLoS ONE*, 16(4): e0251084. <https://doi.org/10.1371/journal.pone.0251084>
- Datong, M. G., Audu, G., & Hilary, U. A. (2017). Logistic Regression Analysis of Risk Factors for Hepatitis B Virus Infection in Jos – Plateau, Nigeria. *International Journal of Innovative Scientific & Engineering Technologies Research*, 5(3):1-5.
- GBD (2016). Causes of Death Collaborators AAAbafati C, Abbas KM, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of

Disease Study 2016. *Lancet* (London, England) 2017;390:1151–210.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied Logistic Regression*. Second edition. John Wiley & Sons Inc: New York.

Mohd S. M., Said, S. M., Lim, P. Y., Azman, A. Z., & Abu Hassan, M. R. (2019). Risk factors for hepatitis C infection among adult patients in Kedah state, Malaysia: A case–control study. *PLoS ONE* 14(10): e0224459. <https://doi.org/10.1371/journal.pone.0224459>.

Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS) report 2018. <http://www.ciheb.org/PHIA>. Accessed 16th October, 2021.

World Health Organization. Global report on hepatitis B 2021. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b>. Accessed 16th October, 2021.

World Health Organization. Global report on hepatitis B 2020. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b>. Accessed 16th October, 2021.

Zhang, H., Li, Q., Sun, J., Wang, C., & Gu, Q. (2011). Seroprevalence and Risk Factors for Hepatitis B Infection in an Adult Population in Northeast China. *International Journal of Medical Sciences*, 8(4), 321-331