Tanzania Journal of Science 50(3): 676-691, 2024
ISSN 0856-1761, e-ISSN 2507-7961

Advanced Principal Component Analysis of Various Risk Factors of Hepatitis B Prevalence in Nigeria.

Nureni O. Adeboye[1*]; Olumide S. Adesina[2]; Habeeb A. Afolabi[1]; Timothy A. Ogunleye[1]: Mutairu K. Kolawole[3]

[1]*Department of Statistics, Faculty of Basic and Applied Science, Osun State University, P.M.B 4494 Osogbo, Nigeria*
[2]Johannesburg Business School, University of Johannesburg, South Africa.
[3]*Department of Mathematics, Faculty of Basic and Applied Science, Osun State University, P.M.B 4494 Osogbo, Nigeria.*
[*]*Corresponding author, e-mail: nureni.adeboye@uniosun.edu.ng +2348033348141*
Received Sep 11 2023, Revised 30 Aug 2024, Accepted 20 Sep 2024, Publ. 30 Sept. 2024
https://dx.doi.org/10.4314/tjs.v50i3.21

Abstract

Hepatitis B virus (HBV) is an infectious disease globally estimated to have caused between 500,000 to 1.2 million deaths annually. HBV prevalence is still high in Nigeria. Thus, this research aimed to identify factors germane to the widespread of HBV infection in an apparent clinical survey. The methods of analysis used were frequency, percentage and Principal component analysis (PCA). This was achieved through hospital record extracts of consultant hepatologists and the dimensionality reduction of the acquired data while retaining essential factors that are germane to the prevalence of HBV infection. The findings revealed that out of seventeen components evaluated in the study, the PCA retained 15 components of which Eigen-values are greater than 1.00. The symptoms retained in every component were fever, muscle pain, fatigue, loss of appetite, blood in vomit, jaundice, pale stool, nausea, blood in faeces, weight loss, malaise, abdominal pain, joint ache, swollen of lower extremities, confusion and yellow eye. The symptoms were listed in accordance with their level of relevance for diagnosing HBV in patients, and all the variables retained accounted for 94.278% variation in the prevalence of HBV infection, with the majority of the infected populace found among the adults (18 -64 years).

Keywords: Eigen-values, Eigen-vectors, Hepatitis B, Principal Component Analysis, Symptoms

Introduction

The widespread of HBV infection varies significantly in different areas. It differs considerably among the European Union, where the prevalence is in the range of 0.1% to 5.5% with Italy and Romania having the highest estimated number of infected cases both above 1 million (Ahmad et al. 2018). The prevalence of HBV infection is lowest in North America, Australia and New Zealand with less than 1%. According to Patel et al. (2019), the prevalence of HBV and HDV in the US among adults between 2011-2016 was 0.36% and 3.4% among non-Hispanic Asians. It is between 2 - 4% in Japan, 5 – 18% in China and the highest rate of 15 – 20% recorded in Taiwan as well as several other countries in South Asia, with Pakistan being the country with one of the highest burdens of chronic hepatitis and mortality globally, due to liver failure and hepatocellular carcinomas (Syed et al. 2009 and Qureshi et al. 2010). Wasley et al. (2010) examined trends in the prevalence of HBV infection in

the US in the wake of widespread hepatitis B vaccination between the periods 1988 – 1994 and 1999 – 2006. The population of ages greater than 6 years were tested for antibodies to HBV core antigen (anti-HBc), HBV surface antigen (HBsAg), and antibody to HBV surface antigen (anti-HBs). Prevalence estimates were weighted and age-adjusted prevalence of anti-HBc (4.7%) and HBsAg (0.27%) of 1999 - 2006 were found not to be statistically different from what they were during 1988–1994 (5.4% and 0.38%, respectively). The authors conjectured that as a result of the widespread vaccination program, the prevalence decreased greatly among US children, but impacted little changes among adults.

The mortality risk of HBV carriers in Africa is estimated at 25% with approximately 50 million chronic carriers (Sonderup et al. 2017). According to World Health Organization (2023), the regional office for Africa reports, Nigeria has a prevalence rate of 8.1% among adults aged 15-64 years. In Africa, including Nigeria, hepatitis is a silent epidemic. More than 90 million people are living with hepatitis in the Region, accounting for 26% of the global total. The country has more than 20 million people living with hepatitis B, C, or both; yet more than 80% of the people who have the disease do not know their status, according to some estimates. Hence, the need to educate the populace about the prevalent factors to drive the quest for treatment agitation.

The transmission of HBV has been traced to many ways, with sexual intercourse and mother-to-child transmission being the most common. Musa et al. (2015) reported that vaccination against the hepatitis B virus in the West African nation of Nigeria is lower than in many sub-Saharan African countries. Chronic hepatitis B infection is a global problem; however, Asia and sub-Saharan Africa are most affected by it. The Hepatitis B status of pregnant women is essential for the effective management of the disease and prevention of mother-to-child transmission (Aba and Aminu 2016, Ayoub and Cohen 2016). Ugbebor et al. (2011) evaluated the prevalence of HBV and HBC infections on pregnant women diagnosed in a Nigerian teaching hospital and conjectured that 720 (12.5%) and 206 (3.6%) out of 5,760 pregnant women included in the study were found to be positive for Serum antibodies to HBV and HBC respectively while 33 (0.57%) were found to have mixed infections of HBV and HBC.

The hepatitis B Virus is a blood-borne virus and is roughly 75-200 times more infectious than HIV (Bowyer and Sim 2011). According to Seeger and Mason (2015), the Dane particle of HBV is a spherical lipid-containing structure of approximately 42 to 47nm. The virion consists of a viral envelope; the nucleocapsid is comprised of 120 dimers of core protein and is covered by a capsid membrane embedded with 3 viral envelope proteins, the large (L), middle (M) and small (S) surface proteins. The partially double-stranded DNA genome consists of a mini-strand, which span the full genome, and a plus-strand of DNA spanning roughly two-third of the genome. Upon infection of the liver cells, the genome is converted to covalently closed circular DNA of which the plus strand is used for the transcription of viral proteins (Bowyer and Sim 2000, Nguyen et al. 2008).

HBV infection is the 10th leading cause of death globally, as a significant number of the chronic carriers later develop liver cirrhosis or hepatocellular carcinoma (HCC) and more than one million have been estimated to die annually from HBV-associated liver disease. HBV infections can also induce cancer in patients and the disease is highly integrated with liver cancer (Ramsey et al. 2019, Fanning et al. 2019, Toh et al. 2013). However, antiviral drugs are available for HBV-infected individuals that may prevent the critical consequences of chronic liver disease, which emphasizes the significance of identifying infected individuals and monitoring the prevalence of the disease.

Globally, programs of HBV vaccine have become a continuous phenomenon in line with the recommendation of the World Health Organization. A large number of individuals with chronic HBV are unaware of their infection status. However, the

implementation of vaccination as a primary preventive measure can greatly reduce the risk of infection. Vaccination efficacy among children has been widely studied, but there remain a large proportion of adult populations who are yet to be vaccinated.

Factors influencing immunologic response to HBV vaccine in adults have been inconsistently examined in existing studies. In this study, we conducted a systematic review and meta-analysis to update and assess a more precise estimation of factors influencing hepatitis B. Identifying these factors on time would create awareness for the would-be carrier to commence the necessary treatment on time before the situation becomes more severe.

Many techniques like Linear Discriminant Analysis, t-Distributed Stochastic Neighbor Embedding, Factor Analysis, and Multidimensional scaling have been used in the literature to provide dimensionality reduction and estimate the prevalence of HBV among adults but they all lacked the ability to exonerate the germane factors from not too important ones before modeling the prevalence of HBV, especially in complex datasets. However, PCA is a widely used dimensionality reduction technique that is particularly valuable in simplifying complex datasets by transforming them into a smaller set of uncorrelated variables known as principal components. This helps to simplify the dataset while retaining most of the original information, and identifies the directions along which the variance of the data is maximized. This is particularly useful in understanding the most significant patterns in the data. Unlike PCA, other techniques like Independent Component Analysis (ICA) focus on statistical independence rather variance, which is more applicable when the goal is to separate mixed signals, such as in blind source separation. More so, PCA is relatively simple and computationally efficient, especially for linear data. It is usually prefer over other techniques because of its balance between simplicity, efficiency, and the ability to maximize variance in the data.

Considering the concomitant loss of lives, cost in medication and loss of productive hours, modeling the prevalence risk of HBV requires renewed commitment from the government, non-governmental organizations and all to fight for the complete eradication of the disease.

Materials and Methods

Source of Data.

The data set used in this article was collected as secondary data from the records of State Hospital Ilaro, Ogun State Nigeria on inbound Hepatitis B patients and it contains information on 200 patients who presented themselves for consultation on HBV-related infections. The symptoms reported by the patients were recorded and information about the same patients was collected after being tested for HBV, for the periods of seven (7) months specifically between 24th March – 13th October 2021. The recorded symptoms as reported by the patients were all compared with the results of the HBV diagnosis, and during these periods, eighteen (18) death casualties were recorded out of the 200 patients monitored during the clinical survey.

Ethical Considerations

The collection of the study data followed ethical processes and was duly approved by the medical director, representing the hospital bioethics committee. Both the approval letter and the coded dataset are available and can be assessed as supplementary information to this article.

Study Variables

The patients involved are between the ages of 4 and 90 years of whom 96 are females and 104 are males with their weight measured between 20 kg and 69kg. Seventeen (17) HBV symptoms were recorded during the clinical survey and these are fever, muscle pain, fatigue, loss of appetite, blood in vomit, jaundice, pale stool, nausea, blood in faeces, weight loss, malaise, abdominal pain, joint ache, swollen of lower extremities, itchy skin, confusion and yellow eye. From the dataset, the ages of the patients are recorded in years while gender was encoded in ordinal form as “2” for Males and “1” for Females. Other features are encoded in integers (“0” for non-

presence and "1" for the presence of the symptoms) while deaths were coded as '0" and '1' for the patients discharged alive.

Methods of Data Analysis.

The analytical technique adopted for this research is an advanced Principal Component Analysis (PCA). It is a data reduction technique that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The following procedures defined the technique of PCA employed in this research:

Given a vector $X' = \{x_1, x_2, x_3, \cdots, x_k\}$ with variance-covariance matrix Σ, eigenvalues $\lambda_1, \lambda_2, \cdots \lambda_k \geq 0$ and liner combinations $Y_j = \alpha_{j1}X_1 + \alpha_{j2}X_2 + \cdots \alpha_{jk}X_k$ of the elements of $X, where\ \alpha_j\ (eigenvector)\ is$ a vector of k components $\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jk}$. Then

$Var(Y_j) = \alpha'_j \sum \alpha_j \quad ; \quad j = 1,2, \cdots, k$ (1)

The linear combinations $Y_1, Y_2, \cdots, Y_k$ are the principal components and are uncorrelated with variances in equation (1) and the first step is to observe a linear combination of $\alpha'_1 X$ with maximum variance, such that

$\alpha'_1 X = \alpha_{11}X_1 + \alpha_{12}X_2 + \cdots + \alpha_{1k}X_k = \sum_{i=1}^{k} \alpha_{1i}X_i$ (2)

When the original variables X_{ji} are substituted into the first PCs, we obtain the new PC observations, Y_{ij} as

$Y_{ij} = \alpha'_{j1}X_{11} + \alpha'_{j2}X_{12} + \cdots + \alpha'_{jk}X_{jk}$ (3)

$i = 1, 2, \cdots, n; j = 1, 2, \cdots, k$

To find the first PC, we seek for a_1, such that

$Y_i = z_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{\rho 1}X_\rho = a_{i1}X$ (4)

$Subject\ to$:

$Var(Y_i) = Var(a'_1 X) = a'_1 a$ is a maximum (i)

$a'_1 a = 1$ (ii)

Note that equation (3) is the general dimensionality equation to obtain a PC while equation (4) is the specific model for obtaining each PC.

To maximize $Var(Y_i)$ subject to $a'_1 a = 1$, we define the Lagrangian function.

$L_{(a_1)} = a'_1 \sum a_1 - \lambda(a'_1 a - 1)$ (5)

Where λ is the Lagrangian Multiplier. To maximize $L_{(a_1)}$, we differentiate $L_{(a_1)}\ w.r.t\ a_1$ and equate to $zero$. Thus,

$$\frac{\partial L(a1)}{\partial a_1} = 2a_1 \sum - 2\lambda a_1 = 0 \quad (6)$$

$(\Sigma - \lambda I)a_1 = 0$ (7)

Where $\lambda = eigen\ value\ of\ \Sigma\ and\ a_1 is\ the\ corresponding\ eigen - vector\ of\ \lambda.$

Ignoring $a_1 = 0$ (negligible), then,

$(\Sigma - \lambda I) = 0$ (8)

And this implies that $\sum = \lambda I$ (9)

To find the second PC, we seek for a_2, such that

$Z_2 = a_{2i}X$

$Subject\ to$:

$Var(Y_i) = Var(a'_2 X) = a'_2 a$ is a maximum (i)

$a_{2i}a_2 = 1$ (ii)

$a_{1i}a_2 = 0$ (iii)

To maximize $Var(Y_i)$ subject to $a'_2 a = 1$, we define the Lagrangian function as

$L(a_2) = a_{21} \sum a_2 - \lambda(a_{2i}a_2 - 1) - \theta a_2^1 a_1$ (10)

Thus;

$$\frac{\partial L(a_2)}{\partial a_2} = 2a_2 \sum - 2\lambda a_2 - \theta_{a1} = 0 \quad (11)$$

$$= 2\,(\sum - \lambda I)\, a_2 - \theta_{a1} = 0 \quad (12)$$

Multiply by a_i to get;

$$2a_{1i} \sum a_2 - 2a_{1i}\lambda a_2 - a_{1i} - a_{1i}\theta_{al} = 0 \quad (13)$$

$$2a_{1i} \sum a_2 - 2a_{1i}\lambda a_2 - a_{1i} - a_{1i}\theta_{a1} = 0 \quad (14)$$

$$2a_{1i} \sum a_2 - 2a_{1i}\lambda a_2 - a_{1i} - a_{1i}\theta_{al} = 0 \quad (15)$$

$$2a_{1i} \sum a_2 - \theta = 0 \quad (16)$$

$$2a_1'' \lambda a_2 - \theta = 0 \quad (17)$$

$$2a_{1i}\lambda a_2 - \theta = 0 \quad (18)$$

$$2(\sum - \lambda I)\, a_2 = 0 \quad (19)$$

$$(\sum - \lambda I)\, a_2 = 0 \quad (20)$$

$$\sum a_2 = \lambda a_2 \quad (21)$$

The same procedures are involved for deriving the subsequent PC. The eigenvector $\alpha_j = \alpha_1, \alpha_2, \ldots, \alpha_p$ is the measure of the importance of a measured variable for a given PC. The proportion of variance tells us the PC that best explained the original variables, and the measure is given as:

$$\varphi_p = \frac{\sum_{j=1}^{q} \lambda j}{P} = \frac{\sum_{j=1}^{q} Var(Z_j)}{P} \quad (22)$$

Results and Discussion

Table 1 presents the socio-demographic and clinical symptoms analyses of patients involved in the research clinical survey. 105(52.5%) were males while the remaining 95(47.5%) were females. 154(77%) were adults, 33(16.5%) were children while 13(6.5%) were older adults. Of the 200 patients involved in the research clinical survey, 182 were discharged while 18 did not survive the treatment after been diagnosed with seventeen (17) different symptoms. Of the 182 patients discharged after receiving treatment, 95 were males while 87 were females. 138 adults who constituted the largest populace of the infected patients were discharged while 16 did not survive the treatment.

Table 1: Socio-demographic and Clinical Characteristics Analysis

Factors	**Frequency**	**Percentage**	**Discharged**	
			YES	**NO**
Gender: Male	105	52.5	95	10
Female	95	47.5	87	08
Total	**200**	**100**	**182**	**18**
Age Group- Older Adults (65 years & Above)	13	6.5	12	01
	154	**77**	138	16
Adults (18 – 64 years)	33	16.5	32	01
Children (Below 18 years)				
Total	**200**	**100**	**182**	**18**
Muscle Pain-	06	03		
YES	194	97	182	18
NO				

Total		**200**	**100**		
Weight	Loss-	09	4.5	182	18
YES		191	95.5		
NO					
Table		**200**	**100**		
Fever-		18	91	182	18
YES		182	09		
NO					
Table		**200**	**100**		
Jaundice-		03	6.5	182	18
YES		187	93.5		
NO					
Total		**200**	**100**		
Abdominal	Pain-	05	2.5	182	18
YES		195	97.5		
NO					
Total		**200**	**100**		
Nausea-		12	94	182	18
YES		188	06		
NO					
Total		**200**	**100**		
Itchy	Skin-	24	12	182	18
YES		176	88		
NO					
Total		**200**	**100**		
Fatigue-		34	17	182	18
YES		166	83		
NO					
Total		**200**	**100**		
Malaise-		194	97	182	18
YES		06	03		
NO					
Total		**200**	**100**		
Joint	Aches-	05	2.5	182	18
YES		195	97.5		
NO					
Total		**200**	**100**		
Pale	Stools-	13	6.5	182	18
YES		187	93.5		
NO					
Total		**200**	**100**		

Confusion-		02	01	182	18
YES		198	99		
NO					
Total		**200**	**100**		
Blood in Vomit-		15	7.5	182	18
YES		185	92.5		
NO					
Total		**200**	**100**		
Blood in Faeces-		10	05	182	18
YES		190	95		
NO					
Total		**200**	**100**		
Swelling of the Lower Extremes-		03	1.5	182	18
YES		197	98.5		
NO					
Total		**200**	**100**		
Yellow Eyes-		02	01	182	18
YES		198	99		
NO					
Total		**200**	**100**		
Loss of Appetite-		22	11	182	18
YES		178	89		
NO					

Table 2 shows the communalities, which is the proportion of each variable's variance that can be explained by the principal components. The initial value of the communality in a principal components analysis is 1 and the values in the extraction column indicate the proportion of each variable's variance that can be explained by the principal components. Symptoms with high values are well represented in the common factor space, while those with low values are not well represented. In this result, there are no particularly low values except that of muscle pain with the lowest extraction value. They are the reproduced variances from the number of components that you have saved. You can find these values on the diagonal of the reproduced correlation matrix.

Table 2: Proportion of Variance of the Hepatitis B

Communalities	**Initial**	**Extraction**
MUSCLE PAIN	1.000	.363
WEIGHT LOSS	1.000	.999
FEVER	1.000	.680
JAUNDICE	1.000	.999

ABDOMINAL PAIN	1.000	.999
NAUSEA	1.000	.999
ITCHY SKIN	1.000	.998
FATIGUE	1.000	.997
MALAISE	1.000	.999
JOINT ACHES	1.000	.999
PALE STOOLS	1.000	.999
CONFUSION	1.000	1.000
BLOOD IN VOMIT	1.000	.999
BLOOD IN FEACES	1.000	.999
SWELLING OF THE LOWER EXTREMITES	1.000	1.000
YELLOW EYES	1.000	1.000
LOSS OF APPETITE	1.000	.998

Table 3 gives the correlations between the original variables. Before conducting a principal components analysis, it is necessary to check the correlations between the variables. If any of the correlations are too high (say above 0.9), one may need to remove one of the variables from the analysis, as the two variables seem to be measuring the same thing. Another alternative would be to combine the variables in some way (perhaps by taking the average). If the correlations are too low, say below 0.1, then one or more of the variables might load only onto one principal component (in other words, make its own principal component). The data for this study showed that there is no form of severe correlation between the variables of Hepatitis B symptoms observed.

Table 3: Correlation Matrix of Hepatitis B Symptoms.

SYMPTOMS	MUSCLE PAIN	WEIGHT LOSS	FEVER	JAUNDICE	ABDOMINAL PAIN	NAUSEA	ITCHY SKIN	FATIGUE	MALAISE	JOINT ACHES	PALE STOOLS	CONFUSION	BLOOD IN VOMIT	BLOOD IN FEACES	SWELLING OF THE LOWER EXTREMITES	YELLOW EYES	LOSS OF APPETITE
1	1																
2	-0.04	1															
3	0.05	-0.07	1														
4	-0.05	-0.06	-0.08	1													
5	-0.03	-0.04	-0.05	-0.04	1												
6	-0.04	-0.06	-0.08	-0.07	-0.04	1											
7	-0.07	-0.08	-0.12	-0.10	-0.06	-0.09	1										
8	-0.08	-0.10	-0.14	-0.12	-0.07	-0.11	-0.17	1									
9	-0.03	-0.04	-0.06	-0.05	-0.03	-0.04	-0.07	-0.08	1								
10	-0.03	-0.04	-0.05	-0.04	-0.03	-0.04	-0.06	-0.07	-0.03	1							
11	-0.05	-0.06	-0.08	-0.07	-0.04	-0.07	-0.10	-0.12	-0.05	-0.04	1						

1	-	-	-	-	-	-	-	-	-	-	-							
2	0.02	0.02	0.03	0.03	0.02	0.03	0.04	0.05	0.02	0.02	0.03							
1	-	-	-	-	-	-	-	-	-	-	-	-	1					
3	0.05	0.06	0.09	0.08	0.05	0.07	0.11	0.13	0.05	0.05	0.08	0.03						
1	-	-	-	-	-	-	-	-	-	-	-	-	-	1				
4	0.04	0.05	0.07	0.06	0.04	0.06	0.09	0.10	0.04	0.04	0.06	0.02	0.07					
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1			
5	0.02	0.03	0.04	0.03	0.02	0.03	0.05	0.06	0.02	0.02	0.03	0.01	0.04	0.03				
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
6	0.02	0.02	0.03	0.03	0.02	0.03	0.04	0.05	0.02	0.02	0.03	0.01	0.03	0.02	0.01			
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	
7	0.06	0.08	0.11	0.09	0.06	0.09	0.13	0.16	0.06	0.06	0.09	0.04	0.10	0.08	0.04	0.04		

Table 4 presents the KMO measures which vary between 0 and 1. This test provides a minimum standard which should be passed before a PCA should be conducted. Values closer to 1 are preferable and a value of 0.6 is a suggested minimum. In this research, a value of 0.8 was obtained and this signifies the sample considered for this study is extremely adequate for a PCA.

Table 4: Kasier-Meyer-Okin Measure of Sampling Adequacy and Bartlett test of Sphericity.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.8
Bartlett's Test of Sphericity	Approx. Chi-Square	623.053
	Degree of freedom	136
	Sig.	.000

The initial number of factors is the same as the number of variables used in the PCA. However, not all the 17 factors will be retained. In this study, majority of the factors were retained due to their Eigen value greater than 1.0 as presented in Table 5. The first 15 components were retained and the remaining two components were discarded as their Eigen values are below 1.0. The retained components had about 94.278% variation of the total variables adopted and this was accomplished in the fifteenth row. This implies that the first fifteen factors together accounted for 94.278% of the total variation in the prevalence of hepatitis B. The first factor accounted for the most variance and hence has the highest eigenvalue, and the next factor accounted for as much of the leftover variance as it can. Hence, each successive factor accounted for lesser variance.

Figure 1 shows the scree plot graphs of the eigenvalue against the component. It was observed that as the component increases the Eigenvalues tends to zero. Only the first-fifteen component has an Eigen-value above 1.00 which is the criteria for retaining a component.

Table 5: Eigen Values of Hepatitis B Symptoms and Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.192	7.011	7.011	1.192	7.011	7.011
2	1.160	6.826	13.837	1.160	6.826	13.837
3	1.130	6.647	20.485	1.130	6.647	20.485
4	1.103	6.488	26.972	1.103	6.488	26.972
5	1.077	6.337	33.309	1.077	6.337	33.309
6	1.070	6.291	39.601	1.070	6.291	39.601
7	1.066	6.268	45.869	1.066	6.268	45.869
8	1.056	6.210	52.079	1.056	6.210	52.079
9	1.049	6.170	58.249	1.049	6.170	58.249
10	1.034	6.082	64.331	1.034	6.082	64.331
11	1.028	6.046	70.378	1.028	6.046	70.378
12	1.026	6.033	76.411	1.026	6.033	76.411
13	1.016	5.978	82.388	1.016	5.978	82.388
14	1.011	5.948	88.336	1.011	5.948	88.336
15	1.010	5.942	94.278	1.010	5.942	94.278
16	.957	5.631	99.909			
17	.015	.091	100.000			

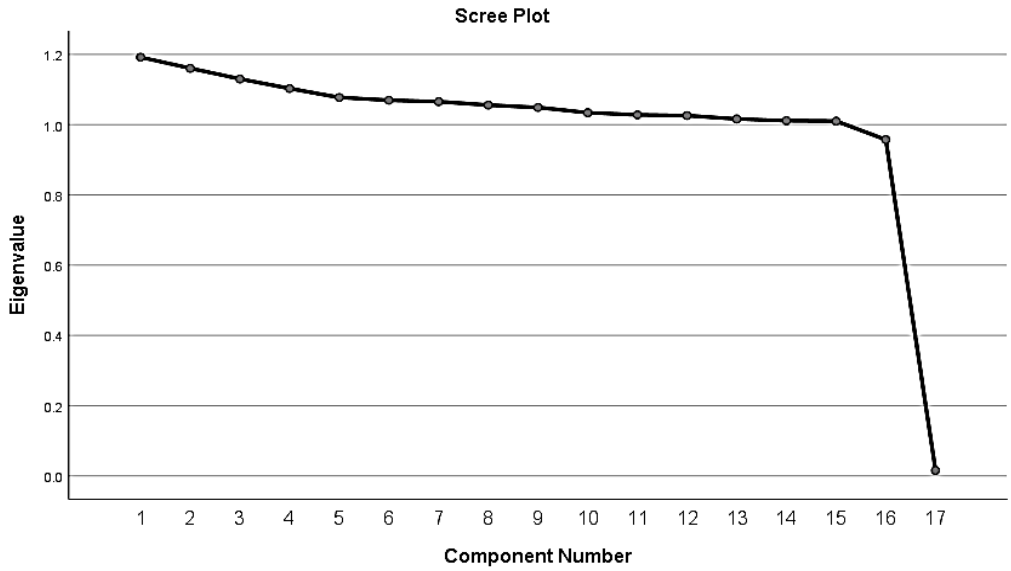

Figure 1: Scree Plot of Eigenvalue against Symptom

Table 6 contains the summary of the non-rotated factor loadings, which are the correlations between the retained components and the factors. Any HBV symptoms below 0.3 on any component are a redundant factor. In the first component, the symptom found to be relevant is fever while muscle pain and fatigue are relevant for detecting the prevalent of HBV in the second component. The third component revealed that only loss of appetite was found to be relevant while only blood in vomit was the retained symptom in the fourth component. In the fifth component, jaundice and pale stool was found to be the relevant symptoms while in the sixth component, pale stool was reconfirmed to be the only relevant symptoms. At the seventh component, nausea was the only symptom while blood in feaces, weight loss, and malaise were the retained symptoms in components eight, nine and ten. In the eleventh and twelfth components, symptoms retained were abdominal pain and joint ache. Swelling of lower extremities was found to be the retained symptoms in the thirteenth component while confusion and yellow eye were found to be the severe symptoms in the fourteenth component. Yellow eye was reaffirmed in the fifteenth component as the only symptoms relevant for detecting the prevalent of HBV in diagnosed patients.

Except for itchy skin, which did not match the criteria of being kept in any of the 15 components, it was discovered that all other 16 symptoms of Hepatitis B were preserved. Thus, the identified prevalent risk factors of HBV in this study are fever, muscle pain, fatigue, jaundice, pale stool, nausea, blood in faeces, weight loss, malaise, abdominal pain, joint ache, swelling of lower extremities, confusion and yellow eye.

Table 6: Component Matrix Showing the Retained Symptoms of HBV

	Component														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Muscle Pain	.29	.51	-.07	-.12	-.03	.00	-.01	-.02	-.02	-.03	-.02	.00	-.02	-.01	.00
Weight Loss	.04	-.06	.03	.14	.08	.00	.07	.39	.90	-.34	-.10	.00	-.05	-.04	.00
Fever	.41	.69	-.09	-.14	-.03	.00	-.02	-.03	-.02	-.03	-.01	.00	-.01	-.01	.00
Jaundice	.06	-.09	.06	.29	.38	-.07	-.39	-.22	-.09	-.12	-.05	.00	-.04	-.03	.00
Abdominal Pain	.03	-.04	.02	.08	.04	.00	.02	.06	.05	.28	.61	.72	-.13	-.08	.00
Nausea	.05	-.08	.05	.23	.21	.00	.85	-.36	-.12	-.12	-.06	.00	-.04	-.03	.00
Itchy Skin	.18	-.45	-.08	-.04	-.07	.00	-.03	-.05	-.03	-.05	-.03	.00	-.02	-.02	.00
Fatigue	-.93	.30	-.08	-.16	-.04	.00	-.02	-.04	-.02	-.04	-.02	.00	-.02	-.02	.00
Malaise	.03	-.04	.02	.09	.04	.00	.03	.80	.70	.85	-.50	.00	-.09	-.06	.00
Joint Aches	.03	-.04	.02	.08	.04	.00	.02	.06	.05	.28	.61	-.72	-.13	-.08	.00
Pale Stools	.06	-.09	.06	.29	.38	.73	-.39	-.22	-.09	-.11	-.05	.00	-.04	-.03	.00
Confusio	.015	-.02	.01	.04	.02	.00	.01	.02	0.30	.06	.05	.00	.13	.69	-.71
Blood in Vomit	.07	-.11	.08	.47	-.84	.00	-.11	-.13	-.06	-.09	-.05	.00	-.03	-.03	.00
Blood In Feaces	.044	-.06	.04	.17	.10	.00	.10	.84	-.44	-.38	-.08	.00	-.05	-.04	.00
Swelling of The Lower Extremites	.02	-.03	.01	.05	.02	.00	.01	.04	.03	.10	.08	.00	.96	-.22	.00
Yellow Eyes	.02	-.02	.01	.04	.02	.00	.01	.03	.02	.06	.05	.00	.13	.69	.71
Loss of Appetite	.14	-.28	.71	-.61	-.09	.00	-.04	-.06	-.03	-.06	-.03	.00	-.03	-.02	.00

Conclusion and Recommendation

This study applied the clinical survey data of patients diagnosed for HBV in a government hospital to explore factors associated with the prevalence of HBV in Nigeria. The findings suggest that fever, muscle pain, fatigue, loss of appetite, blood in vomit, jaundice, pale stool, nausea, blood in faeces, weight loss, malaise, abdominal pain, joint ache, swollen of lower extremities, confusion and yellow eye arranged in chronological order, had high impacts on the prevalence of HBV among the infected patients in Nigeria, mostly adults between the ages of 18 – 64 years who constituted about 77% of the surveyed patients. The findings about the aforementioned age group is in tune with the year 2023 regional office for Africa reports published by the World Health Organization. HBV infection in Sub-Saharan Africa presents a range of risk factors similar to the identified symptoms in this study, which can vary depending on the stage of the disease (acute or chronic). More so, two of the identified risk factors in this study, a yellowing of the skin and eye have been established in literature as the hallmark symptoms of HBV and is prevalent across sub-Saharan Africa. These factors all together, accounted for 94.278% variation in the prevalence of HBV infection in Nigeria. It is pertinent to mention that the clinical survey witnessed 18 deaths among the 200 infected patients while 182 patients were alive and some still undergoing treatments.

Therefore, medical personnel are advised that fever is the first symptom to carefully watch out for in in-patient during Hepatitis B diagnosis closely followed by muscle pain and fatigue which can trigger loss of appetite leading to blood in the patient vomit. The percentage of these symptoms alone contributing to the prevalence of HBV is 26.972% which is approximately ratio 1:4 in diagnosed patients.

Furthermore, the government should entrench continuous funding of the HBV immunization program for all categories of ages while healthcare givers should be given the matching order to always ensure that patients already tested positive for HBV must not be allowed to escalate to a level at which he/she starts to lose blood either through vomit or other forms of excretion, if otherwise, an immediate emergency should be placed on the patient status in order to avoid significant weight loss. The clinical survey deduced that a patient with yellow eye that tends to have seizures must be placed on special HBV treatment even before running other diagnostic tests in other to prevent loss of life for the inbound patient. Lastly, public health initiatives that encourage young adults to seek a prompt and appropriate medical diagnosis of their HBV status are still essential.

Conflict of Interest

The authors declare that no conflicts of interest exist.

Acknowledgements

The authors wish to acknowledge the supports given by the Medical Director of the State Hospital, Ilaro Ogun State Nigeria.

References

Aba HO and Aminu M 2016 Seroprevalence of hepatitis B virus serological markers among pregnant Nigerian women. *Ann. Afr. Med*. 15(1):20-27.

Ahmad AA, Falla AM, Duffell E, Noori T, Bechini A, Reintjes R and Veldhuijzen IK 2018 Estimating the scale of chronic hepatitis B virus infection among migrants in EU/EEA countries. *BMC Infect. Diseases*. 18:34.

Ayoub WS and Cohen E 2016 Hepatitis B Management in the Pregnant Patient: An Update. *J. Clin. Trans. Hepatol.* 4(3): 241–247.

Bowyer SM and Sim JG 2000 Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *J. Gen. Virol*. 81(Pt2):379-92.

Fanning GC, Zoulim F, Hou J and Singapore S 2019 Therapeutic strategies for hepatitis B virus

infection: towards a cure. *Nat. Rev. Drug Discov*. 18: 827–844.

Musa BM, Bussell S, Borodo MM, Samaila AA and Femi OL 2015 Prevalence of hepatitis B virus Infection in Nigeria, 2000-2013: A systematic review and

meta-analysis. *Niger. J. Clin. Pract.* 18 (2): 163-172.

Nguyen DH, Ludgate L and Hu J 2008 Hepatitis B virus-cell interactions and pathogenesis. *J Cell. Physiol*. 216(2):289-294.

Patel EU, Thio CL, Boon D, Thomas DL and Tobian AR 2019 Prevalence of Hepatitis B and Hepatitis D Virus Infections in the United States, 2011–2016, *Clin. Infect. Dis.* 69 (4):709– 712.

Qureshi H, Bile, KM, Jooma R, Alam SE and Afrid HUR 2010 Prevalence of hepatitis B and C viral infections in Pakistan: findings of a national survey appealing for effective prevention and control measures. *EMHJ - East Mediter. Health J.* 16 (Supp.): 15-23.

Ramsey SD, Unger JM, Baker LH, Little RF, Loomba R, Hwang JP, Chugh R, Konerman MA, Arnold K, Menter AR, Thomas E, Michels RM, Jorgensen CW, Burton GV, Bhadkamkar NA and Hershman DL 2019 Prevalence of Hepatitis B Virus, Hepatitis C Virus, and HIV Infection Among Patients with Newly Diagnosed Cancer from Academic and Community Oncology Practices. *JAMA Oncol.* 5(4):497–505.

Seeger C and Mason WS 2015 Molecular biology of hepatitis B virus infection. *Virology* 479-480:672-686.

Sonderup MW, Afihene M, Ally R, Apica B, Awuku Y, Cunha L, Dusheiko G, Gogela N,Lohouès-Kouacou MJ, Lam P, Lesi O, Mbaye PS, Musabeyezu E, Musau B, Ojo O, Rwegasha J, Scholz B, Shewaye AB, Tzeuton C, Kassianides C, Spearman CW 2017 Gastroenterology and Hepatology Association of sub-Saharan Africa (GHASSA). Hepatitis C in sub-Saharan Africa: the current status and recommendations for achieving elimination by 2030. *Lancet Gastroenterol. Hepatol.* 2(12):910-919.

Syed AA, Rafe M.J, Donahue HQ and Sten HV 2009 Hepatitis B and hepatitis C in Pakistan: prevalence and risk factors. *Int. J. Infect. Diseases*, 13 (1): 9-19.

Toh ST, Jin Y, Liu L, Wang J, Babrzadeh F, Gharizadeh B, Ronaghi M, Toh HC, Chow PK, Chung AY, Ooi LL and Lee CG 2013 Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis* 34(4):787-98.

Ugbebor O, Aigbirior M, Osazuwa F, Enabudoso E and Zabayo O 2011 The prevalence of hepatitis B and C viral infections among pregnant women. *N Am J Med Sci*. 3(5):238-241.

WHO 2023 In Nigeria, boosting viral hepatitis awareness and treatment. Retrieved on 5th August 2023 from https://www.afro.who.int/countries/nigeria/news/nigeria.

Wasley A, Kruszon-Moran D, Kuhnert W, Edgar P, Simard LF, McQuillan G and Bell B 2010 The Prevalence of Hepatitis B Virus Infection in the United States in the Era of Vaccination. *J. Infect. Dis.* 202 (2): 192– 201.