Science Journal of Applied Mathematics and Statistics
2023; 11(1): 1-16
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20231101.11
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

On Statistical Analysis of Eight Sexually Transmitted Diseases Using Categorical Analysis of Variance: A Case Study of the University of Nigeria Teaching Hospital

Nnaemeka Martin Eze[1, *], Chimeremeze Davidson Sibigem[1], Oluchukwu Chukwuemeka Asogwa[2], Chinonso Michael Eze[1], Samson Offorma Ugwu[1], Felix Obi Ohanuba[1]

[1]Department of Statistics, University of Nigeria, Nsukka, Nigeria

[2]Department of Mathematics and Statistics, Alex Ekwueme Federal University Ndufu-Alike Ikwo, Nigeria

Email address:

nnaemeka.eze@unn.edu.ng (Nnaemeka Martin Eze)

[*]Corresponding author

To cite this article:

Nnaemeka Martin Eze, Chimeremeze Davidson Sibigem, Oluchukwu Chukwuemeka Asogwa, Chinonso Michael Eze, Samson Offorma Ugwu, Felix Obi Ohanuba. On Statistical Analysis of Eight Sexually Transmitted Diseases using Categorical Analysis of Variance: A Case Study of the University of Nigeria Teaching Hospital. *Science Journal of Applied Mathematics and Statistics.* Vol. 11, No. 1, 2023, pp. 1-16. doi: 10.11648/j.sjams.20231101.11

Received: February 6, 2022; **Accepted**: February 28, 2022; **Published**: January 13, 2023

Abstract: This paper focused on the statistical analysis of eight Sexually Transmitted Infections (STIs) reported in the University of Nigeria Teaching Hospital from 2010-2020. A population of 20,704 patients was recorded to have contracted eight (8) selected STIs. Prevalence analysis was computed to determine the most prevalent STI. Two-way CATANOVA cross-classification was computed to ascertain age group and gender that suffer more of each of these infections. Logistic regression model was fitted to predict reproductive status of patients that suffer the most prevalent STI. The prevalence analysis results showed Gonorrhea infection as the most prevalent STI. Two-way CATANOVA results for Gonorrhea and Chlamydia infections showed that there were significant difference in gender, age and interaction effects, significant difference in age and interaction effect for Trichomoniases infection, significant difference in age for Syphilis and HIV infections but no significant difference in gender, age and interaction effects for Human Papillomavirus (HPV), Hepatitis B Virus and Herpes infections. The results showed that the percentage of male that suffers STIs is more than the percentage of female, the percentage of 30-39 years that suffer STIs is more than the percentage of any other age group and the percentage of people without STIs history is more than the percentage of those that have history of them. Logistic regression results on Gonorrhea infection showed that an increase in age, body mass index, blood pressure, blood sugar, bacteria quantity, and Gonorrhea history were associated with an increased likelihood of the Gonorrhea patient being infertile.

Keywords: Chi-square Test, Contigency Table, Odds Ratio, Significance Level, Prediction

1. Introduction

Nowadays there are several Sexually Transmitted Infections (STIs) in our societies. According to medical experts, these STIs have their origin and the reports showed that some of these infections are incurable. Through the investigation of the U.S. Department of Health and Human Services, it was noted that STIs can be contacted through several ways in which one of them is sexual practices [18]. Sexually Transmitted Infections (STIs) also known as Sexually Transmitted Diseases (STDs) are harmful micro-organisms that is very hard to control its growth in the body of its host. These infections are easily contacted through sex. Most STIs initially do not show symptoms. According to medical experts, infections can be called diseases only when they show symptoms and this is the reason STDs are known as STIs. Medical experts had said that the infections can easily be spread when there is no presence of symptoms of these infections. Some of the symptoms of STIs are vaginal discharge, penile discharge, ulcers on or around the genitals, and pelvic pain. Some STIs may cause infertility in both

male and female and also poor development of a baby if contacted before or during pregnancy. Different bacteria, viruses, fungi, and parasites pathogenic are the major causes of STIs. Some of the bacterial STI are chlamydia infections, gonorrhea or gonococci infection, cancroids, granuloma inguinal, and syphilis. Some of the viral STIs are genital herpes, HIV/AIDS, Viral hepatitis (Hepatitis B virus), and genital warts. Some of the fungal STIs are candidiasis and Parasitic STIs include crab louse, scabies, and Trichomoniasis [18]. Despite the contamination of some STIs through sex, one can contact them through blood and tissues, breastfeeding or during child delivery. The contamination of STIs from one and another or from surrounding objects can be prevented [4]. Azmi et al presented their prevalence analysis from child bearing age women and the result showed that the prevalence of C. trachomatis infection was 0.6% and 0.5%, among symptomatic and asymptomatic women respectively, N. gonorrhoeae was 0.9% and 2.2%, T. pallidum 0.0% and 0.0%, and Tr. vaginalis was 0.7% and 0.5%. It was noted from the results that there was no significant difference in the prevalence rates between symptomatic and asymptomatic women [2]. Kesah et al stated that improvement in hand washing, clean toilets, abstaining from sex, condom usage, rational employment of examination methods, improved medical diagnostics testing both men and women, attitude change and prevention education has to be consistently highlighted [12]. Otaru and Ogbonda studied the application of categorical data-nested design of knowledge and control practices of Hepatitis B Virus (HBV) infection using two-way CATANOVA technique. They considered frequency data from university students in three universities involving response rate of student's knowledge and control practices of HBV infection using a scale of good, fair and poor. The results showed that the responses from students in various faculties in the three considered universities were not significant. They concluded that irrespective of different faculties and universities student's knowledge and control practices of HBV infections was not significantly influenced [16]. Deyheul et al studied infertility rate risk factors and the result showed that the infertility in men and women could be caused by sexually transmitted infections and hormonal disorders. Some lifestyle factors can also cause infertility such as obesity, nutrition, smoking/alcohol consumption, mobile phone use, sexual violence and anxiety [7].

It has been known that Sexually Transmitted Infections (STIs) have sporadically increased over the years and of course have caused more harm than good in our societies. These infections could lead to various dangerous ailments such as infertility, pelvic inflammatory disease in woman, ectopic pregnancy, and serious effect on pregnancy which might lead to miscarriage, failure of development of new baby, blindness, congenital defects and so on.

The aims of this study are to know the most prevalent Sexually Transmitted Infection (STI) among the reported cases of STIs in the University of Nigeria Teaching Hospital; the Gender and Ages that always suffered each of these STIs; and examine the productive status of patients suffering the most prevalent STI.

This study focuses only on eight (8) major sexually transmitted infections (Chlamydia, Gonorrhea, Syphilis, Herpes, Hepatitis B, Trichomoniasis, Human Immunodeficiency Virus (HIV), Human Papillomavirus (HPV)) contracted by both male and female that has attained sexual age as reported at University of Nigeria Teaching Hospital (UNTH) from 2010 to 2020.

The significance of this study tends to educate Nigerians and the world at large about the existence of sexually transmitted infections in our societies and its risk factors. It will also notify people on the most prevalent STI, the gender and age interval that is more likely to be at risk of each of these infections and more precisely, educate them on how to take precautionary measures.

2. Materials and Methods

2.1. Data and Sampling Design

The data used in this study were secondary data collected on eight (8) types of Sexually Transmitted Infections (STIs) reported in the department of Micro Biology, University of Nigeria Teaching Hospital (UNTH). A population of 20,704 patients that reported to have contracted eight (8) selected STIs which include; Chlamydia (4,855), Gonorrhea (6,850), Syphilis (1,680), Trichomoniasis (1,770), Herpes (483), Hepatitis-B (602), Human Papillomavirus (619) and Human Immunodeficiency Virus (3,845) in the years 2010 through 2020 were collected and the prevalence method of analysis was used to ascertain the most prevalent STI among them. Furthermore, the researchers have interest in carrying out more analysis on the most prevalent infection among the eight (8) STIs. In order to determine the most prevalent STI, the record showed that there were 6,850 reported cases of the most prevalent STI and a sample size of 364 was selected from the reported cases of the infection using simple random sample technique. The sample size was obtained by the application of Andrew Fisher's formula at 95% confidence level. This sample size formula that was recommended by [10] is given as:

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} P(1-P)}{d^2} \quad (1)$$

Where n is the desired sample size (if target population is more than 10,000), $Z^2_{1-\frac{\alpha}{2}}$ is the square of the standard normal deviation at the required confidence level of 95% (i.e., at 5% type I error is 1.96), d is the 5% level of statistical significance (i.e., the absolute error or precision which is 0.05), P is the expected proportion in population based on previous studies or pilot studies. Fisher recommended that 50% can be used as the value of the expected proportion P when there is no availability of the previously expected proportion [10].

$$n = \frac{(1.96)^2 \times 0.5 \times (1-0.5)}{0.05^2}$$

$$n = 384.16$$

The population of Gonorrhea patients this study is less than 10,000, therefore, the formula for infinite population was used.

$$nf = \frac{n}{1+\frac{n}{N}} \quad (2)$$

nf is the desired sample size, when the population is less than 10,000, n is the desired sample when population is more than 10,000, N is the estimated population size of reported STIs (6,850).

$$nf = \frac{384.16}{1+\frac{384.16}{6850}} = 364$$

Hence, Sample size is 364 patients.

2.2. Ethical Approval

We addressed ethical issues in this study by making sure that anonymity and confidentiality are highly maintained when the need arises either from the data collection or any sources of information; and the consent of patients was respected. Therefore, all procedures performed in this research that involved patients and healthcare workers were in accordance with the ethical standards of the University of Nigeria Teaching Hospital (UNTH).

2.3. Models

2.3.1. Prevalence Rate

Prevalence is an epidemiology characteristic which is easily measured using survey data or medical records. To establish prevalence, researchers randomly select a sample (smaller group) from the entire population they want to describe. Using random selection methods increases the chances that the characteristics of the sample will be representative of (similar to) the characteristics of the population. For representative sample, prevalence is the number of people in the sample with the characteristics of interest divided by the total number of people in the sample.

$$\left(\text{i.e., Prevalence formula} = \frac{\text{number of people in the sample with the characteristics of interest}}{\text{total number of people in the sample}}\right)$$

2.3.2. Catanova

The categorical analysis of variance (CATANOVA) is a technique designed to identify the variation between treatments of interest to the researcher. This CATANOVA is used to solve problem in analysis of variance when the observations are nominal without any underlying metric and it was also formulated to solve erroneous analysis of nominal data by using chi-square test [15, 16]. In addition, there are several methods for analyzing categorical data in which some of these methods use data transformation before proceeding to analyse the data. The method to be used may depend on the classification of categorical data [9, 11, 14, 20]. In this research, we adopted two-way CATANOVA and we assumed that no loss in generality.

Table 1 shows the data layout for two-way cross classification or a randomized complete block design in which a K-dimensional vector $[n_{ijk}]$ of nominal responses are observed in frequencies in the ijth plot. In this Table 1, the main factor A ranging from 1 to I and main factor B ranging from 1 to J have from 1 to K quanta responses per unit. Table 2 depicted CATANOVA table that contains the source of variation, degrees of freedom (df), sum of squares (SS) which is the trace of its variance-covariance matrix, test ratio from chi-square calculated, critical value from chi-square tabulated and hypotheses for the study.

The assumptions of the categorical data are:

1) Multi-nominal distribution

$$P(\{n_{ijk}\}; \{\pi_{ijk}\}) = \binom{n_{ij}}{n_{ij1},\cdots,n_{ijK}} \prod_{k=1}^{K} (\pi_{ijk})^{n_{ijk}}$$

$$n_{ijk} = 0, 1, \cdots, n_{ij} \text{ and } \pi_{ijk} = \frac{n_{ijk}}{n_{ij}};\ 0 \le \pi_{ijk} \le 1$$

2) Independence: The levels and blocks are each act independently. That is, n_{ijk} and $n_{i'j'k}$ are statistically independent $\forall i \neq i'$ and $\forall j \neq j'$.
3) Constant variance: $\text{var}(n_{ijk}) = n\pi_{ijk}(1-\pi_{ijk})$. The variance is not constant because it depends on i, j and k.
4) $\pi_{ijk} > 0$, $\sum_{k=1}^{K} \pi_{ijk} = 1$, $\sum_{k}^{K} n_{ijk}$ is held fixed (i.e., grand total over k for j)

Table 1. *The data layout for two-way CATANOVA cross-classification or randomized complete block design.*

A (i)	B (j)												
	b1				b2				...	bJ			
	1	2	...	k	1	2	...	k	...	1	2	...	k
1	n_{111}	n_{112}	...	n_{11k}	n_{121}	n_{122}	...	n_{12k}	...	n_{1J1}	n_{1J2}	...	n_{1Jk}
2	n_{211}	n_{212}	...	n_{21k}	n_{221}	n_{222}	...	n_{22k}	...	n_{2J1}	n_{2J2}	...	n_{2Jk}
.	...		...		...		...		...	...		...	
.	...		...		...		...		...	..		...	
.	...		...		...		...			...		...	
I	n_{i11}	n_{i22}	...	n_{i1k}	n_{i21}	n_{i22}	...	n_{i2k}	...	n_{iJ1}	n_{iJ2}	...	n_{iJk}

Table 2. *Summary for two-way catanova cross classification of nominal data.*

source	df	SS	MS	Test Ratio	Critical Value	Hypothesis
Row (Ai)	I-1	RSS	MSA	χ^2_{RT}	$\chi^2_{(I-1)(K-1)}$	$HO_R: \pi_{ijk} = \pi_{jk} \forall_i$
Column (Bj)	J-1	CSS	MSB	χ^2_{CT}	$\chi^2_{(J-1)(K-1)}$	$HO_C: \pi_{ijk} = \pi_{ik} \forall_j$
Interaction (AB)	(I-1)(J-1)	NSS	MSAB	χ^2_{NT}	$\chi^2_{(I-1)(J-1)(K-1)}$	$HO_{RC}: \pi_{ijk} = \pi_{k} \forall_{ij}$

source	df	SS	MS	Test Ratio	Critical Value	Hypothesis
Weight Units	n-IJ	WUSS	UMS	-		-
Total	n-1	TSS	TMS	-		-

Computation of Sum of Squares

Total Sum of Square (TSS) = $n - \frac{\sum_k n^2_{..k}}{n}$

where $n_{..k} = \sum_{ij} n_{ijk}$

Within Unit Sum of Square (WUSS) = $n - \sum_{ij} \frac{\sum_k n^2_{ijk}}{n_{ij}}$

Between Row Sum of Square (BRSS) = $n - \sum_i \frac{\sum_k n^2_{i.k}}{n_i}$

where $n_{i.k} = \sum_j n_{ijk}$

Between Column Sum of Square (BCSS) = $n - \sum_j \frac{\sum_k n^2_{.jk}}{n_j}$

where $n_{.jk} = \sum_i n_{ijk}$

Row Sum of Square (RSS) = TSS – BSS

Column Sum of Square (CSS) = TSS – BCSS

Interaction Sum of Square (NSS) = BCSS + BRSS – TSS – WUSS

Two-way CATANOVA cross classification model is given by

$$E(\pi_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

Where $\hat{\pi}_{ijk}$ is the probability that kth observation occurs in the ith level of factor A and jth level of factor B, i.e., $\hat{\pi}_{ijk} = P_{ijk} = \frac{n_{ijk}}{n_{ij}}$, ($n_{ijk}$ is the kth observation in the ijth cell, n_{ij} is the sum of kth observation in the ijth cells, i.e., $n_{ij} = \sum_k n_{ijk}$), μ is a constant for kth observation, μ is a constant for kth response, α_i (i = 1,2, … , I) is the effect of the ith level of factor A, β_j (j = 1,2, … , J) is the effect of the jth level of factor B, γ_{ij}((i = 1,2, … , I) and (j = 1,2, … , J)) is the effect between the ith level of factor A and jth level of factor B. In nominal data, sum of square is the trace of its variance-covariance matrix and the parameter π_{ijk} may be considered fixed or random with probability density $h(\pi_{ijk})$ ranging from 0 to 1 depending on whether I and J are random or fixed [1, 14, 15, 19].

Hypotheses

$H_{0R}: \pi_{ijk} = \pi_{jk}, i.e\ \alpha_i = 0\ \forall_i$ (There is no row effect)

$H_{1R}: \pi_{ijk} \neq \pi_{jk},, i.e\ \alpha_i \neq 0\ for\ at\ least\ one\ (i)$ (There is row effect)

$H_{0C}: \pi_{ijk} = \pi_{ik}, i.e\ \beta_j = 0\ \forall_j$ (There is no column effect)

$H_{1C}: \pi_{ijk} \neq \pi_{ik}, i.e\ \beta_j \neq 0\ for\ at\ least\ one\ (j)$ (There is column effect)

$H_{0RC}: \pi_{ijk} = \pi_k, \gamma_{ij} = 0 \forall_{ij}$ (There is no interaction effect)

$H_{1RC}: \pi_{ijk} \neq \pi_k, \gamma_{ij} \neq 0\ for\ at\ least\ one\ pair\ (ij)$ (There is an interaction effect)

Test Statistic

$$\chi^2_{RT} = \frac{(K-1)(n-1)RSS}{TSS} \sim \chi^2_{(I-1)(K-1)};\ \alpha$$

$$\chi^2_{CT} = \frac{(K-1)(n-1)CSS}{TSS} \sim \chi^2_{(J-1)(K-1)};\ \alpha$$

$$\chi^2_{NT} = \frac{(K-1)(n-1)NSS}{TSS} \sim \chi^2_{(I-1)(J-1)(K-1)};\ \alpha$$

Decision rule

Reject H_{0R} if $\chi^2_{RT} \geq \chi^2_{(I-1)(K-1)}$, H_{0C} if $\chi^2_{CT} \geq \chi^2_{(J-1)(K-1)}$, and H_{0RC} if $\chi^2_{NT} \geq \chi^2_{(I-1)(J-1)(K-1)}$ at specified level of significance (5%). Otherwise accept.

2.3.3. Logistic Regression

This is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analysis, the logistic regression is a predictive analysis used to describe data and to explain the relationship between one dependent binary response variable, which takes values 1 and 0, and one or more nominal, ordinal, interval or ratio level independent variable (s). The logistic regression gives each predictor a coefficient which measures its independent contribution to variation in the dependent variable. The dependent variable Y takes the value 1 if the response is ‘’yes’’ and takes a value 0 if the response is ‘’no’’. Logistic regression calculates the probability of success over probability of failure. The results of the analysis are in the form of an odds ratio [3].

The model form for predicted probabilities is expressed as a natural logarithm (ln) of the odds ratio.

Logistic model is given by

$$\ln\left[\frac{P(Y)}{1-P(Y)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \quad (3)$$

$$\frac{P(Y)}{1-P(Y)} = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m} \quad (4)$$

$$P(Y) = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m} - P(Y)e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m} \quad (5)$$

$$P(Y) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_m X_m}} \quad (6)$$

Where; $\frac{P(Y)}{1-P(Y)}$ is the odds ratio, $\ln\left[\frac{P(Y)}{1-P(Y)}\right]$ is the log odds or ‘’logit’’ of the outcomes, Y is the dichotomous outcome, P(Y = 1) is the probability of an event, $\times_i$ (i = 1,2, … , m) are the predictors, β_i(i = 0,1,2, … , m) are unknown regression parameters to be estimated and β_0 is the intercept.

Goodness of Fit Test

It is also known as Hosmer-Lemeshow test which represents a Chi-square test used for testing the adequacy of the model for fitting the data. The null hypothesis is that the model is adequate to fit the data and we will only reject this null hypothesis if the p-value is less than 0.05.

It is given as:

$$H = \sum_{i=1}^{g} \frac{(O-E)^2}{E} \quad (7)$$

Where, O and E denote the observed events, and expected events.

Table 3. *Values of the logistic regression model when the independent variable is dichotomous.*

Outcome variable (Y)	Independent variable (X)	
	X = 1	X = 0
Y = 1	$P(1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$P(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
Y = 0	$1 - P(1) = \frac{1}{1+e^{\beta_0+\beta_1}}$	$1 - P(0) = \frac{1}{1+e^{\beta_0}}$

$$Odds\ ratio\ (OR) = \frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} \Big/ \frac{1}{1+e^{\beta_0+\beta_1}}}{\frac{e^{\beta_0}}{1+e^{\beta_0}} \Big/ \frac{1}{1+e^{\beta_0}}} \quad (8)$$

$$= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{(\beta_0+\beta_1)-\beta_0} \quad (9)$$

$$= e^{\beta_1} \quad (10)$$

Hence, for logistic regression with a dichotomous independent variable coded 1 and 0, the relationship between the odds ratio and the regression coefficient is

$$Odds\ ratio(OR) = e^{\beta_1}$$

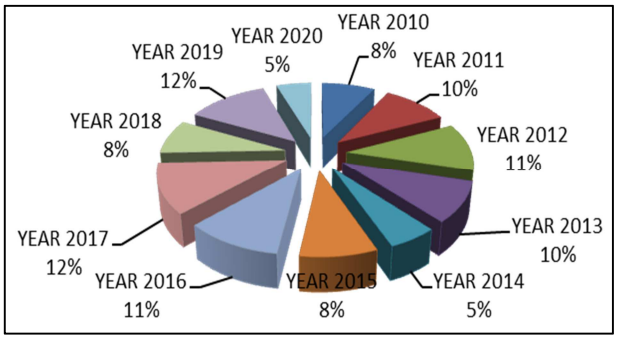

Figure 1. *Yearly percentage of reported cases of Sexually Transmitted Infections.*

Table 4. *Reported cases of selected STIs from 2010-2020.*

YEARS	GON.	CHL.	HIV	TRICHO.	SYPH.	HPV	HEPA. B	HERPES	TOTAL
2010	603	201	414	159	137	18	61	85	1678
2011	941	167	455	173	226	71	28	11	2072
2012	987	437	441	158	148	15	73	35	2294
2013	434	698	397	178	76	88	24	43	1938
2014	392	116	243	147	64	78	35	33	1108
2015	503	576	260	131	115	31	84	15	1715
2016	741	705	272	188	123	23	95	35	2182
2017	982	461	262	118	326	54	159	58	2420
2018	316	516	337	174	178	98	19	16	1654
2019	751	826	418	181	148	119	17	89	2549
2020	200	152	346	163	139	24	7	63	1094
Total	6850	4855	3845	1770	1680	619	602	483	20,704

CHL.= Chlamydia, GON.= Gonorrhea, SYPH.= Syphilis, Herpes, HPV = Human Papillomavirus, TRICHO.=Trichomoniasis, HEPA B.= Hepatitis B Virus, HIV= Human Immunodeficiency Virus

3. Results and Discussion

Figure 1 is a pie chart representation of yearly percentage of reported cases of Sexually Transmitted Infections (STIs) as depicted in Table 4. As can be seen from this Figure 1, the year 2019 (12%) and 2017 (12%) had the highest reported cases of eight (8) types of STIs and this was followed by the year 2016 (11%), 2012 (11%), 2013 (10%), 2011 (10%), 2018 (8%), 2015 (8%), 2010 (8%), 2020 (5%) and 2014 (5%). This Figure 1 shows that there is a difference in yearly report of STIs and this may be due to lack of knowledge about the harmfulness of STIs in the society.

Table 5. *Prevalence rate of the eight (8) selected sexually transmitted infections (2010-2020).*

Infections (STI)	Prevalence Rate	Percentage of Prevalence Rate
CHLAMYDIA	0.2344	23.44
GONORRHEA	0.3308	33.08
SYPHILIS	0.0812	8.12
TRICHOMONIASIS	0.0854	8.54
HERPES	0.0233	2.33
HEPATITIS B	0.0290	2.90
HPV	0.0298	2.98
HIV	0.1856	18.56

CHL.= Chlamydia, GON.= Gonorrhea, SYPH.= Syphilis, Herpes, HPV = Human Papillomavirus, Trichomoniasis, HEPA B.= Hepatitis B Virus, HIV= Human Immunodeficiency Virus

$$\text{Prevalence formula} = \frac{number\ of\ each\ STI}{Total\ number\ of\ STI} \times 100$$

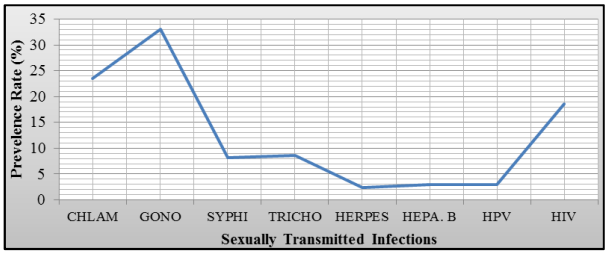

Figure 2. *Prevalence rate of reported cases of sexually transmitted infections.*

Figure 2 is a graphical representation of results of the prevalence rate percentage as depicted in Table 5. As can be seen from this Figure 2, Gonorrhea infection with 33.08% rate appears to be the most prevalent among the eight selected sexually transmitted infections reported in the University of Nigeria Teaching Hospital from 2010-2020 when compared with Chlamydia, Syphilis, Trichomoniasis, Herpes, Human Papilloma Virus (HPV), Hepatitis B, Human Immunodeficiency Virus (HIV) with 23.44%, 8.12%, 8.54%, 2.33%, 2.90%, 2.98% and 18.56% respectively.

3.1. CATANOVA on Gonorrhea, Chlamydia, Syphilis, Herpes, Human Papilloma Virus, Trichomoniasis, Hepatitis B Virus and Human Immunodeficiency Virus Infections

Table 6. *Table for Gonorrhea significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	6.17	18.99	3.841	significant, (reject H_{0R})
Age (Column)	4	29.2	89.89	9.49	significant, (reject H_{0C})
Gender*Age	4	9.8	30.17	9.49	significant, (reject H_{0RC})
Within unit	354	92.34	-	-	
Total	363	117.91	-	-	

From Table 6, it is easy to see that there is statistically significant difference in gender $\left(\chi^2_{RT(cal)} = 18.99 > \chi^2_{RT(tab)} = 3.84\right)$ and age $\left(\chi^2_{CT(cal)} = 89.89 > \chi^2_{CT(tab)} = 9.49\right)$; and there is also statistically significant difference in the interaction effect $\left(\chi^2_{NT(cal)} = 30.17 > \chi^2_{NT(cal)} = 9.49\right)$ at a 5% significance level. Therefore, the null hypotheses (H_{0R}, H_{0C} and H_{0RC}) are rejected and we conclude that there is a particular gender and age interval that suffer more with Gonorrhea infection. The significant in the interaction effect shows that suffering of Gonorrhea infection by the gender depends on their age intervals. From the analysis, we noticed that the percentage of male (58.8%; n = 214) that suffer Gonorrhea infection is greater than female (41.2%; n = 150) and the percentage of 30-39 years (34.3%; n = 125) is greater than 20-29 years (31.6%; n = 115), 40-49 years (16.8%; n = 61), 50 years and above (9.6%; n = 35) and less than 20 years (7.7%; n = 28). Out of 364 sample of Gonorrhea patients, 239 (65.7%) have Gonorrhea infection history while 74 (20.3%) don't have history of it (see Appendix Table 1 for the data and computation).

The results obtained in this study disagreed with the results of other researchers [6, 17, 21]. According to Piszczek et al, the rate of Gonorrhea infection in both Canada and the United States is more common in women with age interval 15-24 and men with interval 20-24 years [17]. In 2012, Creighton stated that the diagnosis of Gonorrhea rates in the UK among adults aged 20-24 years are highest when compared to other age intervals [6] and Word Health Organization, in 2012, stated that an estimated 78 million new cases of Gonorrhea occurred among 15–49 year-olds [21].

Table 7. *Table for Chlamydia significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	23.07	46.64	3.841	significant, (reject H_{0R})
Age (Column)	4	11.09	22.43	9.49	significant, (reject H_{0C})
Gender*Age	4	27.91	56.42	9.49	significant, (reject H_{0RC})
Within unit	4845	2338.96	-	-	
Total	4854	2401.03	-	-	

The analysis in Table 7 showed that there is statistically significant difference in gender $\left(\chi^2_{RT(cal)} = 46.64 > \chi^2_{RT(tab)} = 3.8\right)$ and age $\left(\chi^2_{CT(cal)} = 22.43 > \chi^2_{CT(tab)} = 9.49\right)$; and there is also statistically significant difference in the interaction effect $\left(\chi^2_{NT(cal)} = 56.42 > \chi^2_{NT(tab)} = 9.49\right)$. Therefore, the null hypotheses (H_{0R}, H_{0C} and H_{0RC}) are rejected and we conclude that at a 5% significance level, there is a particular gender and age interval that suffer more with Chlamydia infection and the significant in the interaction effect shows that suffering of Chlamydia infection by the gender depends on their age intervals. The percentage of female (53.2%; n = 2584) that suffer Chlamydia infection is greater than male (46.8%; n = 2271) and the percentage of 50 years and above (27.8%; n = 1348) is greater than 40-49 years (22.4%; n = 1089), 30-39 years (22.2%; n = 1077), 20-29 years (16.4%; n = 795), and less than 20 years (11.2%; n = 546). Out of 4855 reported cases of Chlamydia patients, 2174 (44.8%) have Chlamydia infection history while 2681 (55.2%) don't have history of it (see Appendix Table 2 for the data).

Table 8. *Table for Trichomoniasis significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	0.23	0.52	3.841	not significant, (accept H_{0R})
Age (Column)	4	51.49	116.39	9.49	significant, (reject H_{0C})
Gender*Age	4	9.22	20.84	9.49	significant, (reject H_{0RC})
Within unit	1760	721.68	-	-	
Total	1769	782.62	-	-	

From the results in Table 8, we noticed that there is no statistically significant difference in gender $\left(\chi^2_{RT(cal)} = 0.52 < \chi^2_{RT(tab)} = 3.84\right)$ but there is a statistically significant difference in age $\left(\chi^2_{CT(cal)} = 116.39 > \chi^2_{CT(tab)} = 9.49\right)$. The no significant difference in gender means that no particular gender suffers more of Trichomoniasis infection than another gender. The significant difference in age means that a particular age group is the most likely age group that suffers from Trichomoniasis infection. It was noticed also that there is a statistically significant difference in the interaction between gender and age at a 5% significance level $\left(\chi^2_{NT(cal)} = 20.84 > \chi^2_{NT(tab)} = 9.49\right)$. The significant difference in the interaction between gender and age groups of Trichomoniasis patients means that the rate of contracting Trichomoniasis infection by males and females differs from one age group to another. Out of 1770 reported cases of Trichomoniasis, 930 (52.5%) are males while 840 (47.5%) are females. The percentage of 30-39 years (34.6%; n = 612) that suffer Trichomoniasis infection is greater than less than 20 years (26.7%; n = 473), 20-29 years (20.2%; n = 357), 40-49 years (9.3%; n = 164) and 50 years and above (9.3%; n = 164). We noticed that 584 (33.0%) have Trichomoniasis infection history while 1186 (67%) don't have history of it (see Appendix Table 3 for the data).

Table 9. *Table for Syphilis significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	0.16	0.39	3.841	not significant, (accept H_{0R})
Age (Column)	4	21.38	52.28	9.49	significant, (reject H_{0C})
Gender*Age	4	1.71	4.18	9.49	not significant, (accept H_{0RC})
Within unit	1670	663.32	-	-	
Total	1679	686.57	-	-	

The results in Table 9 showed that there is no statistically significant difference in gender $\left(\chi^2_{\mathrm{RT(cal)}} = 0.39 < \chi^2_{\mathrm{RT(cal)}} = 3.84\right)$ but there is a statistically significant difference in age $\left(\chi^2_{\mathrm{CT(cal)}} = 52.28 > \chi^2_{\mathrm{CT(tab)}} = 9.49\right)$. The no significant difference in gender means that no particular gender suffers Syphilis infection more than the other. The significant difference in age means that a particular age group is the most likely age group that suffers from Syphilis infection. It was noticed also that there is no statistically significant difference in the interaction between gender and age at a 5% significance level $\left(\chi^2_{\mathrm{NT(cal)}} = 4.18 < \chi^2_{\mathrm{NT(tab)}} = 9.49\right)$. The no significant difference in the interaction between gender and age groups of Syphilis patients means that the rate of contracting Syphilis infection by males and females is the same. Out of 1680 reported cases of Syphilis infection, 846 (50.4%) are males whi le 292 (17.4%) are females. The percentage of 30-39 years, (36.3%; n = 610), that suffer Syphilis infection is greater than 20-29 years (23.7%; n = 398), less than 20 years (18.0%; n = 302), 40-49 years (13.5%; n = 226) and 50 years and above (8.6%; n = 144). Out of 1680 Syphilis patients, 481 (28.6%) have Syphilis infection history while 1199 (71%) don't have history of it (see Appendix Table 4 for the data).

Table 10. *Table for Human Immunodeficiency Virus (HIV) significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	1.64	3.39	3.841	not significant, (accept H_{0R})
Age (Column)	4	27.32	56.37	9.49	significant, (reject H_{0C})
Gender*Age	4	3.87	7.98	9.49	not significant, (accept H_{0RC})
Within unit	3835	1830	-	-	
Total	3844	1862.83	-	-	

The analysis in Table 10 showed that there is no statistically significant difference in gender $\left(\chi^2_{RT(cal)} = 3.39 < \chi^2_{RT(tab)} = 3.84\right)$ but there is statistically significant difference in age groups $\left(\chi^2_{\mathrm{CT(cal)}} = 56.37 > \chi^2_{\mathrm{CT(tab)}} = 9.49\right)$. Moreover, there is no statistically significant difference in the interaction effect since $\chi^2_{NT(cal)} = 7.98 < \chi^2_{\mathrm{NT(tab)}} = 9.49$. It was concluded that at a 5% significance level, there is no particular gender that suffers HIV infection more than the other but there is a particular age group that suffers it more than other age groups. There is no significant difference in the interaction effect shows that suffering of HIV infection by the gender depends on their age intervals. Out of 3845 reported cases of HIV infection, 2040 (53.1%) are males while 1805 (46.9%) are females. The percentage of 30-39 years (30.5%; n = 1172) that suffer HIV infection is greater than 40-49 years (23.5%; n = 904), 50 years and above (20.5%; n = 787), 20-29 years (16.9%; n = 650) and less than 20 years (8.6%; n = 332). Out of 3845 HIV patients, 1585 (41.2%) have HIV infection history while 2260 (58.8%) don't have history of it. (see Appendix Table 5 for the data).

Table 11. *Table for Human Papillomavirus (HPV) significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	0.08	0.17	3.841	not significant, (accept H_{0R})
Age (Column)	4	1.82	3.88	9.49	not significant, (accept H_{0C})

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender*Age	4	0.44	0.92	9.49	not significant, (accept H_{0RC})
Within unit	609	288.74	-	-	
Total	618	291.08	-	-	

The results in Table 11 showed that there is no statistically significant difference in gender ($\chi^2_{RT(cal)} = 0.17 < \chi^2_{(tab)} = 3.84$), there is no statistically significant difference in age ($\chi^2_{CT(cal)} = 3.88 < \chi^2_{(tab)} = 9.49$) and interaction between gender and age ($\chi^2_{NT(cal)} = 0.92 < \chi^2_{NT(tab)} = 9.49$). The no significant difference in gender and age means that no particular gender and age suffer Human Papillomavirus infection more than the other at a 5% significance level. The no significant difference in the interaction between gender and age groups of Human Papillomavirus patients means that the rate of contracting Human Papillomavirus infection by males and females is the same at a 5% significance level. Out of 619 reported cases of HIV infection, 316 (51.1%) are males while 303 (48.9%) are females. 81 (13.1%) are less than 20 years old, 140 (22.6%) are at the age of 20-29 years, 149 (24.1%) are at the age of 30-39 years, 144 (23.3%) are at the age of 40-49 years and 105 (17.0%) are at the age of 50 years and above. We noticed that 234 (37.8%) have HPV infection history while 385 (62.2%) don't have history of it (see Appendix Table 6 for the data).

Table 12. *Table for Hepatitis B Virus significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	0.01	0.03	3.841	not significant, (accept H_{0R})
Age (Column)	4	2.93	6.32	9.49	not significant, (accept H_{0C})
Gender*Age	4	0.49	1.06	9.49	not significant, (accept H_{0RC})
Within unit	592	274.68	-	-	
Total	601	278.11	-	-	

From Table 12, we noticed that there are no statistically significant difference in gender ($\chi^2_{RT(cal)} = 0.03 < \chi^2_{RT(tab)} = 3.84$), age ($\chi^2_{CT(cal)} = 6.32 < \chi^2_{CT(tab)} = 9.49$) and interaction between gender and age ($\chi^2_{NT(cal)} = 1.06 < \chi^2_{NT(tab)} = 9.49$). Therefore, the null hypotheses (H_{0R}, H_{0C} and H_{0RC}) are accepted and we concluded that there is no particular gender or age that suffers Hepatitis B virus infection more than the other at a 5% significance level. Moreover, the no significant difference in the interaction between gender and age groups of Hepatitis B virus patients means that the rate of contracting Hepatitis B virus infection by males and females is the same at a 5% significance level. Out of 602 reported cases of Hepatitis B virus infection, 312 (51.8%) are males while 290 (48.2%) are females. 71 (11.8%) are less than 20 years old, 127 (21.1%) are at the age of 20-29 years, 156 (25.9%) are at the age of 30-39 years, 136 (22.6%) are at the age of 40-49 years and 112 (18.6%) are at the age of 50 years and above. We noticed that 218 (36.2%) have Hepatitis B virus infection history while 384 (63.8%) don't have history of it (see Appendix Table 7 for the data).

Table 13. *Table for Herpes significance of effects.*

Source	DF	Sum of Squares	Test Ratio	Critical Value	Decision
Gender (Row)	1	0.31	0.70	3.841	not significant, (accept H_{0R})
Age (Column)	4	2.73	6.14	9.49	not significant, (accept H_{0C})
Gender*Age	4	0.36	0.82	9.49	not significant, (accept H_{0RC})
Within unit	473	211.26	-	-	
Total	482	214.67	-	-	

The results in Table 13 showed that there are no statistically significant difference in gender ($\chi^2_{RT(Cal)} = 0.70 < \chi^2_{RT(tab)} = 3.84$), age ($\chi^2_{CT(Cal)} = 6.14 < \chi^2_{CT(tab)} = 9.49$) and interaction between gender and age ($\chi^2_{NT(Cal)} = 0.82 < \chi^2_{NT(tab)} = 9.49$). Therefore, the null hypotheses (H_{0R}, H_{0C} and H_{0RC}) are accepted and it was concluded that there is no particular gender or age that suffers Herpes infection more than the other at a 5% significance level. Moreover, the no significant difference in the interaction between gender and age groups of Herpes patients means that the rate of contracting Herpes infection by males and females is the same at a 5% significance level. Out of 483 reported cases of Hepatitis B virus infection, 256 (53.0%) are males while 227 (47.0%) are females. The analysis showed 39 (8.1%) are less than 20 years old, 92 (19.0%) are at the age of 20-29 years, 126 (26.1%) are at the age of 30-39 years, 135 (28.0%) are at the age of 40-49 years and 91 (18.8%) are at the age of 50 years and above. We noticed that 161 (33.3%) have Hepatitis B virus infection history while 322 (66.7%) don't have history of it (see Appendix Table 8 for the data).

3.2. Logistic Regression on Gonorrhea Infection as the Most Prevalent Infection Among the Eight Selected STIs

Table 14. *Logistic regression analysis code sheet for dependent and independent variables data.*

Variable	Description	Codes/ Values	Name	Data type
$\times_1$	Age	Years	Age	Numerical
$\times_2$	History of Gonorrhea	0 = No; 1 = Yes	History	Nominal
$\times_3$	Body Mass Index	kg/m^2	BMI	Numerical
$\times_4$	Blood Pressure	mm Hg	BP	Numerical
$\times_5$	Blood Sugar	mg/dl	BS	Numerical
$\times_6$	Bacteria Quantity	(cfu/ml)*10^8	BQ	Numerical
Y	Dependent variable	0 = fertile; 1 = infertile	Fertility	Nominal

The way a particular data is presented goes a long way in determining its analytical case. In order to prevent some problems usually encountered in poor presentation of data, an extra care is taken, in Table 14, to present the independent variables and their data type and values.

Table 15. *Testing for Multi-collinearity.*

Model	Unstandardized Coefficients		Standardized Coefficients	t	P-value	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	-0.999	0.0207		-4.825	0.000		
AGE	0.011	0.003	0.263	3.590	0.000	0.409	2.447
History of Gonorrhea	0.220	0.086	0.172	2.543	0.012	0.988	1.012
BMI (Kg/m^2)	0.013	0.005	0.114	2.339	0.020	0.926	1.080
BP (mmHg)	0.002	0.001	0.062	1.273	0.047	0.912	1.096
BS (mg/dl)	0.002	0.001	0.140	1.980	0.048	0.436	2.292
BQ (cfu/ml)*10^8	0.131	0.059	0.106	2.218	0.027	0.964	1.037

Multi-collinearity occurs when independent variables in a model are correlated. In logistic regression, this kind of correlation is a problem because independent variables should be independent, i.e., there should be weak or no relationship among themselves. If there is presence of multicollinearity, logistic regression estimates will be unstable and have high standard errors. A researcher can use Tolerance method or Variance Inflation Factor (VIF) method to check this multi-collinearity. Tolerance is used as an indicator of multi-collinearity. The high value of tolerance is an indication that there is no multi-collinearity in the model while the low value of tolerance is known to affect adversely the results associated with the model. The minimum tolerance value should be < 0.25. Variance Inflation Factor (VIF) is the reciprocal of tolerance. It identifies correlation between independent variables and the strength of that correlation. The minimum value of VIF is 1 and has no upper limit. The value between 1 and 4 indicates that there is no correlation between this independent variable and any other variable and it suggests absence of multi-collinearity [8]. From Table 15, the independent variables had no multi-collinearity. Since the Tolerance for the variables were greater than 0.25. Also, to confirm our claim the VIF values were between 1 and 4.

Table 16. *Omnibus Tests of Model Coefficients.*

	Chi-Square	df	P-value
Step	88.217	6	0.000
Block	88.217	6	0.000
Model	88.217	6	0.000

Omnibus Tests of Model Coefficients are used to assess the fitness of the overall logistic regression model. The overall model contains all the considered independent variables, unlike the null model which contains no independent variables. From Table 16, the Omnibus Tests of Model Coefficients tested the model fit to predict the reproductive status (i.e., fertility or infertility) of Gonorrhea patients. It tested the significance of the independent variables coded as age, history, blood sugar, bacteria quantity, body mass index, and blood pressure as predictors of the model with reproductive status as a dependent variable (fertile = 0 and infertile = 1). Also, the results show in Table 16, a chi-square value of 88.217 with 6 degrees of freedom (df) and P-value less than 0.05 (i.e., $\chi_{(6)} = 88.217$, P-value < 0.05). It means that the overall model is statistically significant, that is, the model as a whole fits significantly to predict the reproductive status of Gonorrhea patients better than a model with no predictors at a 5% significance level.

Table 17. *Model summary.*

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	295.170	0.215	0.330

The Cox & Snell and Nagelkerke R^2 seen in Table 17 is similar to R^2 we use in linear regression which gives us an idea of how much of variance in the dependent variable is explained by the independent variables. From the results in Table 17, we observed that 33% (ranging between 21.5% - 33%) of the variance in the outcome variable is being affected by the predictor variables. It is preferable to report the Nagelkerke R^2 because it is a modification of Cox & Snell R^2 that cannot achieve a value of 1but Nagelkerke R^2 can reach a maximum of 1. It can be seen from Nagelkerke's R^2 result that 33.0% of the variance in the outcome variable is affected by predictor variable and it can be said that there is evidence to say that the logistic model is adequate or a good fit for the data.

Table 18. *Hosmer and Lemeshow test.*

Step	Chi-square	Df	P-value
1	5.474	8	0.706

The Hosmer and Lemeshow Test examines if the logistic model fit perfectly with observed group memberships. A Chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. To fit a logistic regression model, the Hosmer and Lemeshow Test must obtained p-value greater than significance level of 0.05. However, we can see from Table 18 that the p-value (P > 0.706) is greater than significance level of 0.05 and this shows that a logistic model can be fitted using the dependent and independent variables.

Table 19. *Classification Table.*

Observed		Predicted		
		Reproductive Status		Percentage Correct
		FERTILE	INFERTILE	
Reproductive Status	FERTILE	276	8	97.2
	INFERTILE	15	65	81.3
Overall Percentage				93.7

From the Table 19, the Gonorrhea patients with predicted probabilities of fertility greater than or equal to 0.5 are classified into the fertile group while those with predicted probabilities of infertility greater than or equal to 0.5 are classified into the infertile group. The model correctly classified 65 (81.3%) Gonorrhea patients into the infertile group; this is known as the sensitivity of prediction, that is, the percentage of occurrences correctly predicted. The model also correctly classified 276 (97.2%) Gonorrhea patients into the fertile group and this is known as a specificity of prediction, that is, the percentage of nonoccurrences correctly predicted. The overall correct prediction was 341 out of 364 Gonorrhea patients with an overall success rate of 93.7%. The model predicted the total number of infertility as 73 against the actual observation of 80 Gonorrhea patients. It predicted 8 (11.0%) infertile Gonorrhea patients that were wrongly classified into the fertile group at the time of actual observation recording and this is known as a false positive prediction. Also, the model predicted the total number of fertility as 291 against the actual observation of 284 Gonorrhea patients. It predicted 15 (5.2%) fertile Gonorrhea patients that were wrongly classified into the infertile group and this is known as a false negative of prediction.

Table 20. *The logistic regression model table to predict infertility level of Gonorrhea patients.*

	B	S. E.	Wald	Df	p-value	Exp (B)	95% C.I. for EXP (B)	
							Lower	Upper
AGE	0.085	0.022	15.241	1	0.000	1.088	1.043	1.135
HISTORY (1)	1.313	0.475	7.645	1	0.006	3.718	1.466	9.431
BMI	0.105	0.042	6.253	1	0.012	1.111	1.023	1.206
BP	0.013	0.010	1.562	1	0.046	1.013	1.002	1.034
BS	0.012	0.006	4.209	1	0.040	1.012	1.001	1.024
BQ	0.846	0.427	3.927	1	0.048	2.331	1.009	5.382
Constant	-11.208	1.830	37.510	1	0.000	0.000		

From Table 20, it was noticed that column 2 shows results for logistic regression model, column 7 shows results for odd ratios related to each independent variable and column 8 shows 95% confidence interval for odd ratios. Thus, the higher these odd ratios being above 1, the more likely a patient is to be infertile. For instance, a patient is likely to be infertile if his/her age increase 1.08 times. This shows that odd ratio gives the magnitude of effect that each independent may have on predicting the outcomes. Moreover, odd ratio (OR) can be less than 1 (< 1), greater than 1 (>1) or equal to 1 (= 1). If odd ratio is 1, then there is no change in odd. If it is less than 1, then odds are decreasing for every unit change in predictor variable. If it is greater than 1, then odds are increasing for every unit change in the predictor variable. The null hypothesis states that OR is equal to 1 (i.e., Ho: OR =1) and if this null hypothesis falls between the upper and lower limit of confidence interval, then we cannot reject the null hypothesis and say that the predictor is not having a significant impact on the odds target group but the predictor is significant if it falls outside the upper and lower limits of the confidence interval. In this study, null hypothesis will be accepted if the p-value is greater or equal to significance level of 0.05 $(\mathrm{i.e.}, \mathrm{p-value} \geq 0.05)$ and be rejected if otherwise. From the results in column 6 of this Table 19, the p-values for Age, History, BMI, BP, BS and BQ are 0.001, 0.006, 0.012, 0.046, 0.040 and 0.048 respectively and they are less than significance level of 0.05 $(i.e., p-value < 0.05)$ and this means that the predictors are significant.

Obtained Model: The following is the fitted logistic regression model

$$\mathrm{Y} = -11.208 + 0.085(\mathrm{Age}) + 1.313(\mathrm{History}) + 0.105(\mathrm{BMI}) + 0.013(\mathrm{BP}) + 0.012(\mathrm{BS}) + 0.846(\mathrm{BQ}) \quad (11)$$

Predictions

Equation (11) is substituted into equation (12) to obtain the general form of the logistic regression equation for prediction.

$$P(Y) = \frac{e^{-11.208+0.085(\text{Age})+1.313(\text{History})+0.105(\text{BMI})+0.013(\text{BP})+0.012(\text{BS})+0.846(\text{BQ})}}{1+e^{-11.208+0.085(\text{Age})+1.313(\text{History})+0.105(\text{BMI})+0.013(\text{BP})+0.012(\text{BS})+0.846(\text{BQ})}} \quad (12)$$

By using equation (12), predictions were made using the following variables:

$$\text{g}\,(Age = 52, History = 1, BMI = 21.333, BP = 130, BS = 279, BQ = 0.103)$$

$$-11.208 + 0.085(52) + 1.313(1) + 0.105(21.333) + 0.013(130) + 0.012(279) + 0.846(0.103) = 1.8901$$

$$P(Y) = \frac{e^{1.8901}}{1+e^{1.8901}} = 0.8688$$

$$\text{g}\,(Age = 25, History = 0, BMI = 26.9, BP = 120, BS = 114, BQ = 0.302)$$

$$-11.208 + 0.085(25) + 1.313(0) + 0.105(26.9) + 0.013(120) + 0.012(114) + 0.846(0.302) = -3.0750$$

$$P(Y) = \frac{e^{-3.0750}}{1+e^{-3.0750}} = 0.0442$$

Table 21. *Probability Computations for Classification of Fertility Status.*

Age	History of Gonorrhea	Body Mass Index (BMI)	Blood Pressure (BP)	Blood Sugar (BS)	Bacteria Quantity (BQ)	Probability (Y)	Classification of Reproductive Status of Gonorrhea Patients
52	1	21.333	130	279	0.103	0.87***	Infertile
25	0	26.9	120	114	0.302	0.04**	Fertile
34	1	26.439	100	168	0.206	0.32**	Fertile
64	1	20.08	130	116	0.502	0.76***	Infertile
40	0	22.676	120	127	0.168	0.10**	Fertile
49	1	26.8	139	132	0.123	0.64***	Infertile
16	0	24.9	120	104	0.033	0.01**	Fertile
38	1	21.4	135	164	0.092	0.35**	Fertile

***P(Y) greater than 0.5 = Infertile
**P(Y) less than 0.5 = Fertile
*P(Y) equal to 0.5 = Equal chances of being Infertile or Fertile

4. Conclusion

The data collected on the eight (8) types of sexually transmitted infections (STIs) recorded from 2010 through 2020 in the department of Micro Biology, University of Nigeria Teaching Hospital were used. Firstly, prevalence analysis method was used to determine the most prevalent sexually transmitted infection among the eight (8) selected infections (Chlamydia, Gonorrhea, Syphilis, Trichomoniasis, Hepatitis B, Herpes, Human papilomavirus (HPV) and Human Immunodeficiency Virus (HIV)) and the result showed that Gonorrhea is the most prevalent STI by 33.08%. Secondly, two-way CATANOVA cross classification was used to ascertain the gender and ages that always suffer with each of these selected infections.

From the Gonorrhea infection results, we noticed that there were statistically significant difference in gender, age and interaction effects between gender and age at 5% level of significance. Our findings depicted that the percentage of males that suffer Gonorrhea infection is more than the percentage of females. The percentage of 30-39 years old that suffer Gonorrhea infection is more than the percentage of any other age intervals. From the findings, we noticed that the rate of suffering Gonorrhea infection depends on the age intervals. The findings also showed that the percentage of people with Gonorrhea history contact the infection more than the percentage of people without history of it.

The results for Chlamydia infection showed that there were statistically significant difference in gender, age and interaction effects between gender and age at a 5% significance level. Our findings showed that the percentage of females that suffer Chlamydia infection is more than the percentage of males and the percentage of 50 and above years old that suffer the Chlamydia is more than the percentage of any other age groups. The finding showed that the rate of suffering Chlamydia infection by the gender depends on their age intervals. The findings showed that Chlamydia infection spread rapidly as the percentage of people without its history suffer it more than the percentage of those with its history.

The results for Trichomoniasis infection showed no statistically significant difference in gender but there were statistically significant difference in age and interaction between gender and age at a 5% significance level. Our findings showed that the suffering of Trichomoniasis infection by males and females are the same. We also noticed that the percentage of 30-39 years old that suffer Trichomoniasis infection is more than the percentage of any other age groups. The finding also showed that the rate of suffering Trichomoniasis infection by the genders is dependent on their age intervals and the spread of it is high given that the percentage of people without Trichomoniasis infection history is more than the percentage of those with its history.

The results for Syphilis infection showed that there were no statistically significant difference in gender and interaction

effects between gender and age at a 5% significance level but there was statistically significant difference in age. Our findings showed that no gender suffers Syphilis infection more than the other. We also noticed that the percentage of 30-39 years old that suffer Syphilis infection is more than the percentage of any other age groups and the rate at which it spreads across the age groups is the same. The findings showed that the spread of Syphilis infection is high seeing as the percentage of people without its history suffers it more than the percentage of those that have history of it.

The results for Human Immunodeficiency Virus (HIV) infection showed that there are no statistically significant difference in gender and interaction effects between gender and age at a 5% significance level but there was statistically significant difference in age. Our findings showed that no gender suffers HIV infection more than the other. We also noticed that the percentage of 30-39 years old that suffer HIV infection is more than the percentage of any other age groups and the rate at which it spreads across the age groups is the same. The findings showed that the HIV infection is spreading rapidly since the percentage of people without its history is more than the percentage of those that have history of it.

The results for Human Papillomavirus (HPV), Hepatitis B Virus and Herpes infections showed that there were no statistically significant difference in gender, age and interaction effects between gender and age at a 5% significance level. Our findings showed that no gender and age groups suffer HPV, Hepatitis B Virus and Herpes infections more than the other and the rates at which each age group suffers these infections are the same. The findings showed that these infections are spreading rapidly since the percentage of people without their history is more than the percentage of those that have their history.

Our findings showed that Gonorrhea infection is the most prevalent STIs. Thus, we were motivated to fit a logistic regression model that will assist the medical practitioners to make a prediction about reproductive status of Gonorrhea patients.

We were motivated to fit a logistic regression model on Gonorrhea infection since it is the most prevalent STI among the eight (8) selected STIs in this study. A logistic regression was performed to determine the effects of the variables coded as age, history, blood sugar, bacteria quantity, body mass index, and blood pressure on the likelihood that a Gonorrhea patient is infertile. The logistic regression model was statistically significant, $\chi_{(6)} = 88.217, \mathrm{P} - \mathrm{value} < 0.001$. The model explained 33.0% (Nagelkerke R^2) of the variance in reproductive status (fertility or infertility) of Gonorrhea patients and correctly classified 93.7% of cases into the fertile and infertile groups. A Gonorrhea patient with a Gonorrhea history is 3.718 times more likely to be infertile than a Gonorrhea patient without a Gonorrhea history. An increase in age, body mass index, blood pressure, blood sugar, and bacteria quantity of a Gonorrhea patient were associated with an increased likelihood of being infertile.

Generally, this statistical model can also be fitted for other STIs using the factors associated with the bacteria that cause them.

This study is an eye opener to different types of sexually transmitted infections for Nigerians and other people in different countries. According to the findings of this study, significant steps should be taken to create awareness and motivate adults the need for regular health check-up for proper termination or cure of these infections. More precisely, the concerned authorities need to make effort to educate people on STIs and this may be through mass-media, social media, schools and any other means of communication. The authorities should also provide appropriate health care facilities in both urban and rural areas with government intervention so that the poor ones can benefit and with these measures, STIs especially Gonorrhea infection with their risk factors will reduce drastically in Nigeria and the world at large.

Data availability Statement

We used secondary data from the department of Micro Biology, University of Nigeria Teaching Hospital. The data had been presented in this research work and any other information needed on the data used in this work will be made available.

Conflict of Interest

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to acknowledge the department of Micro Biology, University of Nigeria Teaching Hospital (UNTH) who made their data available for free for this research.

Appendix

Table A1. *Two-way contingency table depicting response of gender and ages of gonorrhea patients reported in UNTH from 2010-2020.*

Gender (i)	**Age (j)**								
	< 20 (j_1)			**20-29 (j_2)**			**30-39 (j_3)**		
	Gonorrhea History response			**Gonorrhea History response**			**Gonorrhea History response**		
	YES	**NO**	**Total n_{i1}**	**YES**	**NO**	**Total n_{i2}**	**YES**	**NO**	**Total n_{i3}**
Male	2	3	5	30	12	52	87	3	90
Female	4	19	23	53	20	73	33	2	35
Total n_{jk}	6	22	28	83	32	115	120	5	125

Table A1. *Continued.*

Gender (i)	Age (j)						Total $n_{i\,k}$		Total n_i
	40-49 (j_4)			50+ (j_5)					
	Gonorrhea History response			Gonorrhea History response					
	YES	NO	Total n_{i4}	YES	NO	Total n_{i5}	YES	NO	
Male	50	3	53	18	6	24	187	27	214
Female	6	2	8	7	4	11	103	47	150
Total n_{jk}	56	5	61	25	10	35	290	74	364

Computation of Sum of Squares for Gonorrhea Infection

Sample size used is 364 out of 6850 Gonorrhea patients reported cases

$$TSS = 364 - \frac{290^2+74^2}{364} = 117.91$$

$$WUSS = 364 - \frac{2^2+3^2}{5} + \frac{30^2+12^2}{52} + \cdots + \frac{7^2+4^2}{11} = 92.34$$

$$BRSS = 364 - \frac{187^2+27^2}{214} + \frac{103^2+47^2}{105} = 111.74$$

$$BCSS = 364 - \frac{6^2+22^2}{28} + \frac{83^2+32^2}{115} + \cdots + \frac{25^2+10^2}{35} = 88.71$$

$$RSS = 117.91 - 111.74 \ = 6.17$$

$$CSS = 117.91 - 88.71 = 29.2$$

$$NSS = 88.71 + 111.74 - 117.91 - 92.34 = 9.8$$

Chi-square calculated

$$\chi^2{}_{RT} = \frac{(2-1)\times(364-1)\times 6.17}{117.91} = 18.99$$

$$\chi^2{}_{CT} = \frac{(2-1)\times(364-1)\times 29.2}{117.91} = 89.89$$

$$\chi^2{}_{NT} = \frac{(2-1)\times(364-1)\times 9.8}{117.91} = 30.17$$

Chi-square tabulated

$$\chi^2{}_{RT} = \chi^2{}_{(2-1)(2-1)} = \chi^2{}_{(1)}(at\ 5\%\ from\ chi-square\ table = 3.841)$$

$$\chi^2{}_{CT} = \chi^2{}_{(5-1)(2-1)} = \chi^2{}_{(4)}(at\ 5\%\ from\ chi-square\ table = 9.49)$$

$$\chi^2{}_{NT} = \chi^2{}_{(2-1)(5-1)(2-1)} = \chi^2{}_{(4)}(at\ 5\%\ from\ chi-square\ table = 9.49)$$

Table A2. *Two-way contingency table depicting response of gender and ages of Chlamydia patients reported in UNTH from 2010-2020.*

Gender (i)	Age (j)								
	< 20 (j_1)			20-29 (j_2)			30-39 (j_3)		
	Chlamydia History response			Chlamydia History response			Chlamydia History response		
	YES	NO	Total n_{i1}	YES	NO	Total n_{i2}	YES	NO	Total n_{i3}
Male	147	87	234	190	155	345	208	303	511
Female	109	203	312	208	242	450	225	341	566
Total n_{jk}	256	290	546	398	397	795	433	644	1077

Table A2. *Continued.*

Gender (i)	Age (j)						Total $n_{i\,k}$		Total n_i
	40-49 (j_4)			50+ (j_5)					
	Chlamydia History response			Chlamydia History response					
	YES	NO	Total n_{i4}	YES	NO	Total n_{i5}	YES	NO	
Male	306	220	526	284	371	655	1135	1136	2271
Female	202	361	563	295	398	693	1039	1545	2584
Total n_{jk}	508	581	1089	579	769	1348	2174	2681	4855

Table A3. *Two-way contingency table depicting response of gender and ages of Trichomoniasis patients reported in UNTH from 2010-2020.*

Gender (i)	**Age (j)**								
	< 20 (j_1)			**20-29 (j_2)**			**30-39 (j_3)**		
	Trichomoniasis History response			**Trichomoniasis History response**			**Trichomoniasis History response**		
	YES	**NO**	**Total n_{i1}**	**YES**	**NO**	**Total n_{i2}**	**YES**	**NO**	**Total n_{i3}**
Male	132	101	233	70	144	214	51	247	298
Female	111	129	240	34	109	143	83	231	314
Total n_{jk}	243	230	473	104	253	357	134	478	612

Table A3. *Continued.*

Gender (i)	**Age (j)**						**Total $n_{i\,k}$**		**Total n_i**
	40-49 (j_4)			**50+ (j_5)**					
	Trichomoniasis History response			**Trichomoniasis History response**					
	YES	**NO**	**Total n_{i4}**	**YES**	**NO**	**Total n_{i5}**	**YES**	**NO**	
Male	42	50	92	19	74	93	314	616	930
Female	21	51	72	21	50	71	270	570	840
Total n_{jk}	63	101	164	40	124	164	584	1186	1770

Table A4. *Two-way contingency table depicting response of gender and ages of Syphilis patients reported in UNTH from 2010-2020.*

Gender (i)	**Age (j)**								
	< 20 (j_1)			**20-29 (j_2)**			**30-39 (j_3)**		
	Syphilis History response			**Syphilis History response**			**Syphilis History response**		
	YES	**NO**	**Total n_{i1}**	**YES**	**NO**	**Total n_{i2}**	**YES**	**NO**	**Total n_{i3}**
Male	35	92	127	99	127	226	66	252	318
Female	41	134	175	62	110	172	71	221	292
Total $\mathbf{n_{jk}}$	76	226	302	161	237	398	137	473	610

Table A4. *Continued.*

Gender (i)	**Age (j)**						**Total $n_{i\,k}$**		**Total n_i**
	40-49 (j_4)			**50+ (j_5)**					
	Syphilis History response			**Syphilis History response**					
	YES	**NO**	**Total n_{i4}**	**YES**	**NO**	**Total n_{i5}**	**YES**	**NO**	
Male	33	58	91	15	69	84	248	598	846
Female	47	88	135	12	48	60	233	601	834
Total n_{jk}	80	146	226	27	117	144	481	1199	1680

Table A5. *Two-way contingency table depicting response of gender and ages of HIV patients reported in UNTH from 2010-2020.*

Gender (i)	**Age (j)**								
	< 20 (j_1)			**20-29 (j_2)**			**30-39 (j_3)**		
	HIV History response			**HIV History response**			**HIV History response**		
	YES	**NO**	**Total n_{i1}**	**YES**	**NO**	**Total n_{i2}**	**YES**	**NO**	**Total n_{i3}**
Male	63	90	153	123	216	339	242	387	629
Female	57	122	179	112	199	311	179	364	543
Total n_{jk}	120	212	332	235	415	650	421	751	1172

Table A5. *Continued.*

Gender (i)	**Age (j)**						**Total $n_{i\,k}$**		**Total n_i**
	40-49 (j_4)			**50+ (j_5)**					
	HIV History response			**HIV History response**					
	YES	**NO**	**Total n_{i4}**	**YES**	**NO**	**Total n_{i5}**	**YES**	**NO**	
Male	253	282	535	188	196	384	869	1171	2040
Female	193	176	369	175	228	403	716	1089	1805
Total n_{jk}	446	458	904	363	424	787	1585	2260	3845

Table A6. *Two-way contingency table depicting response of gender and ages of Humanpapilloma virus (HPV) patients reported in UNTH from 2010-2020.*

Gender (i)	**Age (j)**								
	< 20 (j_1)			**20-29 (j_2)**			**30-39 (j_3)**		
	HPV History response			**HPV History response**			**HPV History response**		
	YES	**NO**	**Total n_{i1}**	**YES**	**NO**	**Total n_{i2}**	**YES**	**NO**	**Total n_{i3}**
Male	13	26	39	26	47	73	31	42	73
Female	15	27	42	29	38	67	30	46	76
Total n_{jk}	28	53	81	55	85	140	61	88	149

Table A6. *Continued.*

Gender (i)	Age (j)						Total $n_{i\,k}$		Total n_i
	40-49 (j_4)			50+ (j_5)					
	HPV History response			HPV History response					
	YES	NO	Total n_{i4}	YES	NO	Total n_{i5}	YES	NO	
Male	30	44	74	17	40	57	117	199	316
Female	28	42	70	15	33	48	117	186	303
Total n_{jk}	58	86	144	32	73	105	234	385	619

Table A7. *Two-way contingency table depicting response of gender and ages of Hepatitis B Virus (HEPA B) patients reported in UNTH from 2010-2020.*

Gender (i)	Age (j)								
	< 20 (j_1)			20-29 (j_2)			30-39 (j_3)		
	HEPA B History response			HEPA B History response			HEPA B History response		
	YES	NO	Total n_{i1}	YES	NO	Total n_{i2}	YES	NO	Total n_{i3}
Male	8	25	33	26	41	67	32	45	77
Female	11	27	38	23	37	60	29	50	79
Total n_{jk}	19	52	71	49	78	127	61	95	156

Table A7. *Continued.*

Gender (i)	Age (j)						Total $n_{i\,k}$		Total n_i
	40-49 (j_4)			50+ (j_5)					
	HEPA B History response			HEPA B History response					
	YES	NO	Total n_{i4}	YES	NO	Total n_{i5}	YES	NO	
Male	28	46	74	18	43	61	112	200	312
Female	27	35	62	16	35	51	106	184	290
Total n_{jk}	55	81	136	34	78	112	218	384	602

Table A8. *Two-way contingency table depicting response of gender and ages of Herpes patients reported in UNTH from 2010-2020.*

Gender (i)	Age (j)								
	< 20 (j_1)			20-29 (j_2)			30-39 (j_3)		
	Herpes History response			Herpes History response			Herpes History response		
	YES	NO	Total n_{i1}	YES	NO	Total n_{i2}	YES	NO	Total n_{i3}
Male	3	14	17	15	38	53	27	43	70
Female	6	16	22	13	26	39	24	32	56
Total n_{jk}	9	30	39	28	64	92	51	75	126

Table A8. *Continued.*

Gender (i)	Age (j)						Total $n_{i\,k}$		Total n_i
	40-49 (j_4)			50+ (j_5)					
	Herpes History response			Herpes History response					
	YES	NO	Total n_{i4}	YES	NO	Total n_{i5}	YES	NO	
Male	22	47	69	14	33	47	81	175	256
Female	25	41	66	12	32	44	80	147	227
Total n_{jk}	47	88	135	26	65	91	161	322	483

References

[1] Anderson, T. W. (1958). *An introduction to multivariate analysis*. John Wiley: New York.

[2] Azmi M., Muataz A., Ali M. A. and Mohammad S. E (2008). Prevalence of Sexually Transmitted Infections Among Sexually Active Jordanian Females. *Sex Transm. Dis.*, 35 (6): 607-710.

[3] Boateng, E. Y. and Abaye, D. A. (2019). A Review of the Logistic Regression Model with Emphasis on Medical Research. *Journal of Data Analysis and Information Processing*, 7: 190-207.

[4] Centers for Diseases Control and Prevention (2013). Incidence, prevalence and cost of living in the United States.

[5] Centers for Diseases Control and Prevention (2022). Gonorrhea-CDC Fact Sheet. Accessed: 06 January, 2022. Available: https://www.cdc.gov/std/gonorrhea/stdfact-gonorrhea.htm

[6] Creighton, S. (2014). Gonorrhea. *BMJ clinical evidence*, 1604. Accessed: 20 August, 2021. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3931440/

[7] Deyheul N., Mohamaddoost T. and Hossini M. (2017). Infertility related risk factors: A systematic review. *International Journal of women health and reproduction science*, 5 (1): 24-29.

[8] Eze, N. M., Asogwa, O. C. and Eze, C. M. (2021). Principal component factor analysis of some development factors in southern Nigeria and its extension to regression analysis. *Journal of Advances in Mathematics and Computer Science, 36* (3): 132-160. DOI: 10.9734/JAMCS/2021/v36i330351

[9] Fienberg, S. E. (1973). Analysis of incomplete multiway contingency tables. Biometrics, 28: 177-202. DOI: https://doi.org/10.2307/2528967

[10] Fisher, L. D (1998). Self-designing clinical trials. Statistics in Medicine; 17: 1551-1562.

[11] Florian, T. J. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang, 59* (4): 434-446.

[12] Kesah F. N. C., Vincent K. P and Augustine A. (2013). Prevalence and etiology of Sexually Transmitted Infections I gynecologic unit of a developing country. *Annals of tropical medicine and public health, 6* (5): 526.

[13] Mayo Foundation for Medical Education and Research (2022). Gonorrhea. Accessed: 04 January, 2022. Available: https://www.mayoclinic.org/diseases-conditions/gonorrhea/symptoms-causes/syc-20351774

[14] Onukogu, I. B. (2014). Analysis of variance of categorical data-nested designs. *Journal of Statistics: Advances in Theory and Applications*, 12: 109-116.

[15] Onukogu, I. B. (1985). Reasoning by analogy from ANOVA to CATANOVA, *Biom J,* 27: 839-849.

[16] Otaru O. P. and Ogbonda N. P (2020). CATANOVA analysis of knowledge and control practices of hepatitis B virus infection amongst tertiary university students. Galician Medical Journal, 27 (1).

[17] Piszczek, J., St. Jean, R. and Khaliq, Y. (2015). Gonorrhea: Treatment update for an increasingly resistant organism, 148 (2): 82-89. doi: 10.1177/1715163515570111

[18] Scatterwhite, C. L., Torrone, E., Meites, E., Dunne, E. F., Mahajan, R., Ocfemia, C. Su, J., Xu, F. and Weinstock, H. (2013). Sexually transmitted infections among U.S. women and men: Prevalence and incidence estimates, 2008 *Sexually Transmitted Diseases*; 40 (3): 187-193.

[19] Scheffe, H. (1959). *The Analysis of Variance*. Wiley: New York.

[20] Singh, B. (2004). CATANOVA for analysis of nominal data from repeated measures design. *J Ind Soc Agril Statist, 58* (3): 257-268.

[21] World Health Organization (2016). *WHO Guidelines for the Treatment of Neisseria Gonorrhoeae*. Accessed: 12 December, 2021. Available: https://apps.who.int/iris/bitstream/handle/10665/246114/9789241549691-eng.pdf