Asian Journal of
Science, Technology, Engineering, and Art

e-ISSN: 3025-4507
p-ISSN: 3025-5287

Index: Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/AJSTEA.v4i3.9196

Estimation of Binary Logistic Regression Using Three Links Function (Logit, Probit, and Complementary Log Log) in Assessing the Factor That Influence HIV

Tugga H. A.[1], Ogunmola A. O.[2], Bamigbala O.A.[3], Ahmad S.S.[4]
[1,2,3]Federal University Wukari, Taraba State, Nigeria
[4]Federal University Dutsin-Ma, Katsina State, Nigeria
hussainatugga@fuwukari.edu.ng

Article Info:

Submitted:	Revised:	Accepted:	Published:
Feb 13, 2026	May 11, 2026	May 23, 2026	May 28, 2026

Abstract

Human Immunodeficiency Virus (HIV) remains a major global public health concern, with sub-Saharan Africa accounting for a substantial proportion of the global burden of infection. In Nigeria, the HIV epidemic shows geographic and demographic variation shaped by age, sex, socioeconomic status, risk behaviors, and access to healthcare services. Understanding the determinants of HIV infection is therefore essential for effective prevention, early detection, and policy formulation. This study aimed to identify significant demographic determinants of HIV infection and determine the best-fitting binary response model among patients tested at General Hospital Takum, Taraba State, Nigeria, between 2018 and 2023. Binary logistic regression models with logit, probit, and complementary log–log link functions were applied to assess the effects of age, sex, and year on HIV infection status. Model performance was evaluated using goodness-of-fit statistics, including deviance, Pearson chi-square, and Hosmer–Lemeshow tests, as well as model selection criteria based on the Akaike Information Criterion and Bayesian Information Criterion. The results indicate a consistent decline in HIV odds across the study years, significantly higher odds among females, and substantially increased odds among adults aged 30–49 years and those aged 50 years and above. Among the three models, the complementary log–log link function demonstrated the best overall fit, with the lowest AIC and BIC values and non-significant goodness-of-fit tests. The study concludes that age, sex, and year are significant predictors of HIV infection, and that the complementary log–log model provides the most reliable framework for predicting HIV status in this population. These findings contribute to epidemiological modelling by supporting more appropriate link-function selection and offer practical implications for localized HIV prevention strategies in Taraba State, Nigeria.

Keywords: Binary Logistic Regression; Complementary Log–Log Model; HIV Infection; Model Selection; Nigeria

Introduction

Human Immunodeficiency Virus (HIV) remains a major public health challenge globally, with profound implications for health systems, economic productivity, and social welfare. Over the past three decades, HIV has caused more than 25 million deaths worldwide, with Sub-Saharan Africa bearing the highest burden of the epidemic. As of 2014, approximately 36.9 million people were living with HIV globally, of which about 66% resided in Sub-Saharan Africa (Awofala & Ogundele, 2018). In Nigeria, HIV prevalence varies considerably across geopolitical zones and demographic groups. Odimegwu et al. (2017) reported higher infection rates among women and substantial regional heterogeneity, with the South–South region recording the highest prevalence. Persistent challenges such as low risk perception, multiple concurrent sexual partnerships, and inadequate access to healthcare services and HIV related stigma continue to undermine effective prevention and treatment strategies. Heterosexual transmission accounts for over 80% of new HIV infections in Nigeria, followed by mother-to-child transmission (Awoleye & Thron, 2015). Although key populations such as female sex workers, injecting drug users, and men who have sex with men constitute a small proportion of the population, they contribute disproportionately to new infections. Obeagu and Obeagu (2022) emphasized that Nigeria's large untreated HIV population poses a serious challenge to epidemic control, underscoring the importance of early detection and effective antiretroviral therapy coverage.

Identifying demographic determinants of HIV infection is essential for targeted interventions and evidence-based policymaking. Among the available statistical techniques, binary logistics regression is widely used for modeling dichotomous outcomes such as HIV status (positive or negative) (Kudakwashe & Yesuf, 2014).Classical logistic regression is typically estimated using the logit link function. However alternative link functions like probit and complementary log-log can also be applied. Especially in situation when data possess asymmetry or when certain assumptions of the logit model are not fully met. Comparing these link functions can provide deeper insights into model fit, predictive performance, and the nature of underlying relationships. This study therefore estimates and compares logit, probit, and complementary log–log models to assess the effects of age, sex, and year on HIV infection using data from General Hospital Takum, Taraba State.

Literature Review

Several studies have employed regression based approaches to investigate HIV related outcomes in Nigeria and other low income settings. Rimamkyanten et al. (2023) used log-linear models to examine factors associated with HIV/AIDS progression in Adamawa State, finding significant effects of age, gender, marital status, and viral load. Onemayin et al. (2019) applied a k-order binary logistic regression model to assess factors influencing HIV counseling and testing utilization in South-West Nigeria, reporting age and education related disparities. Advanced probit based approaches have also been used in health studies. Seyoum (2018) employed a bivariate probit model to examine correlated health outcomes, while Ochalek et al. (2017) used a recursive bivariate probit model to estimate the causal impact of HIV on employment status in Uganda. These studies highlight the flexibility of alternative link functions in capturing complex relationships. Despite this growing body of literature, most HIV studies in Nigeria rely exclusively on the logit link and focus on national or regional data, with limited attention to local-level analysis. Moreover, few studies formally compare multiple link functions to identify the most appropriate model. This study addresses these gaps by conducting a localized analysis and systematically comparing logit, probit, and complementary log–log models.

Materials and Methods

Study Design and Data Source

This study adopts a retrospective design using secondary data obtained from Honorable Haruna Tsokwa Memorial Hospital (General Hospital), Takum Local Government Area, Taraba State Nigeria. The dataset comprises HIV test records of patients who underwent CD4 count assessment and antiretroviral therapy monitoring between 2018 and 2023.

Variables

The response variable is HIV status, coded as binary (1 = HIV positive, 0 = HIV negative). The explanatory variables include sex (male, female), age group (0–14, 15–29, 30–49, 50+), and year of test (2018–2023).

Methodology

In this section we discuss the binary regression model with it various link function (logit, probit and the complementary log log link functions) with each of these models uses a different functional form to relate the probability p to the regressors, and they are all estimated using maximum likelihood methods (Guris & Caglyan2011). Binary Regression Model provides the basis of multifaceted models for ordinal, nominal and count model. Binary response variable has two categories, one represent the occurrence and other nonoccurrence of an event. Binary Regression Model explores how regressors' affects the chance of occurrence of an event. (Kamal and Pervaiz, 2011).

Binary Logistic Regression Model

Logistic regression models are nonlinear regression model and also a special case of GLMs with binomial random component and a logit link Function. Logistic regressions models are also called logit models which assume a linear relationship with the explanatory variable. The range of the predicted probabilities lies between 0 and 1 (Alan, 2007). For binary response variable Y and explanatory variable X, let

$$\pi(x) = P(Y = 1 / X = x) = 1 - P(Y = 0 / X = x)$$

Multiple Logistic Regressions

This logistic regression model extends to model with multiple explanatory variables

The model for $\pi(x) = \mathrm{P}(y = 1)$ at $\mathrm{X} = (x_{1,\ldots\ldots} x_p)$ of p predictor is expressed as

$$\log it\left[\pi(x)\right] = \log\left[\frac{\pi(xi)}{1 - \pi(xi)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots.. + \beta_p x_p \qquad (1)$$

Where $\pi(x)$ is the probability of having the disease. (Success)

$(1 - \pi(x))$ Is the probability of not having the disease.

Xi is the explanatory variable or the predictors (age, sex and year)

α Gives the odds of having the disease at the baseline i.e. when x=0
β_i Show how the odds differ across the explanatory variable.

The alternative formula, directly specifying the probability of success i.e. having the disease $\pi(x)$ is

$$\pi(x) = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+.....+\beta_p x_p)}}{1+e^{(\alpha+\beta_1 x_1+\beta_2 x_2+.....+\beta_p x_p)}} \tag{2}$$

Conversely the probability of not having the disease is

$$1-\pi(x) = \frac{1}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+.......+\beta_p x_p)}} \tag{3}$$

Probit Model

Another nonlinear model use to analyze the relationship between a binary response and multiple explanatory variables is the probit model (Ruspriyanty and Sofro 2018). In order to explain the behavior of a dichotomous dependent variable the probit model uses the cumulative distribution function (CDF). The model derived from the normal cumulative distribution function (CDF) is referred to as the probit model or the normit model. For binary response having the form

$\pi(x)$ = f(x) for some CDF. Let $\phi(.)$ denote the standard cdf of the class such as the N (0, 1) CDF. The probit model is express as

$$\pi(x) = \phi(\alpha+\beta x) = \phi^{-1}\left[\pi(x)\right] = \alpha+\beta x \tag{4}$$

(Alan 2002).

The probit link function is $\varphi^{-1}(.)$.

Complementary Log-Log Model

The inverse of the logit model create another Binary Regression Model called Complementary log-log Regression. The range of predicted probabilities lies between 0 and 1. The Complementary log-log Regression Model is expressed as;

$$p(y=1/X=x) = 1-\exp\left\{-\exp\beta_p(x_p)\right\} \tag{5}$$

Taking log of the negative log of the complement of $\pi(x) = p(y = 1 / X = x)$

$$\log\{-\log(1-p)\} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_p x_p \qquad (6)$$

The Complementary log-log Model follows Gompertz Distribution. The left hand side of equation (6) is called Complementary log log Link. (Kamal & Pervaiz, 2011). The link for this GLM is called the complementary log-log link, since the log-log link applies to the complement of $\pi(x)$.

Inference for Logistic Regression

To test the hypothesis of the model the procedure adapted was to fit an intercept only model (null model) and then fit a model with more parameters (fitted model). The likelihood ratio test is use to test the significance of the hypothesis:

In this study the following hypothesis will be tested

Null Hypothesis (H_0):

1. The demographic and clinical variables (such as age, sex and year of test) have significant influence on HIV infection.
2. There is no significant difference in model performance among the three link functions.

Alternative Hypothesis (H_1):

1. The demographic and clinical variables significantly influence HIV infection.
2. There is significant difference in model performance among the three link functions.

Likelihood Ratio Test

In order detect lack of fit for a binary response the likelihood ratio statistic to test is often use to test whether certain model parameter are zero by comparing the log likelihood L_1 for fitted model M_1 with L_0 for simpler model M_0. Denote this statistic for testing M_0, given M_1 holds by $G^2(M_0 / M_1)$ the goodness of fit statistic $G^2(M)$ is a special case in which $M_0 = M_1$ and M_1 is the saturated model. In testing whether M fits, we test whether all parameter under the saturated model are zero. The degree of freedom is the difference in number of parameter in the two model, which is the number of parameter in the saturated minus the number of parameter in the assume model. The likelihood ratio statistic for comparing M_1 and M_0 is

$$G^2(M_0/M_1) = G^2(M_0) - G^2(M_1). \tag{1.3}$$

In logistic regression models Parameter estimators have large-sample normal distributions. Thus, inference can use the Wald, likelihood-ratio, score test. For large samples, the three tests usually give similar results. (Agresti. 2002). This study uses the The likelihood-ratio test.

Goodness of Fit Test

It is also called Hosmer-Lemeshow test which represents a χ^2 (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 there are sufficiently strong grounds to do so (traditionally if the p-value is less than 0.05) (Qais. M Abdulqader, 2017). And also in assessing the quality of the fitted models, Pearson's chi-square statistic and the likelihood ratio statistics (G2) which are based on the comparison of the fitted and observed counts were used. If the chi-square is significant, the variable is considered to be significant in the model. (Kudakwashe &Yesuf 2014).

Binary Logistic Regression Assumption

1. The response variable can only take two possible outcomes.
2. There should not be multi-collinearity between the explanatory variables.
3. There should not be extreme outlier.
4. It assumes a linear relationship between the log odds and the explanatory variable.
5. The response variable for this model are mutually independent

Linear Approximation Interpretation for β_i

The S- shape for $\pi(x)$ is determined by the rate of increase or decrease of parameter β in equations (1) and (2). The sign of β indicates whether the curve ascends ($\beta > 0$) i.e $\pi(x)$ increases as x increase or descends ($\beta < 0$), i.e $\pi(x)$ decrease as x increases and the rate of change increases as $|\beta|$ increases. The right-hand side of equation (1) simplifies to a constant When $\beta = 0$. Indicating that $\pi(x)$ is identical at all x, so the curve becomes a horizontal straight line. The binary response Y is then independent of X. (Alan 2007)

The Odds Ratio Interpretation

A vital interpretation of logistic regression uses the odds and odds ratio. Exponentiating both side of eqn (2) the odds of response are (i.e the odds of success)

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha+\beta x) = e^{\alpha}(e^{\beta})^{x} \quad (1.4)$$

Equation 1.4 provides an interpretation for β: the odds multiple by e^{β} for every one unit increase in x. that is, the odds at level x+1 equal the odds at level x multiple by e^{β}. When β=0, $e^{\beta}=1$ (Alan Agresti, 2007).

Results and Discussion

To identify the appropriate link functions which fit the response variables three goodness of fit techniques were adapted to assessed the adequacy of these models on the data presented in the table 1 below

Table 1: Goodness of fit tests for Binary Logistic Regression Using Various Link Functions

Link Function	Test	Chi-Square	P-Value
Logit	Deviance	6066.35	0.958
	Pearson	6421.07	0.073
	Hosmer-Lemeshow	11.48	0.075
Probit	Deviance	6075.31	0.950
	Pearson	6495.27	0.018
	Hosmer-Lemeshow	13.19	0.040
Complementary Log-Log	Deviance	6063.23	0.960
	Pearson	6293.81	0.372
	Hosmer-Lemeshow	6.59	0.361

The goodness-of-fit results in Table 1 shows that the logit model fits the data well, with acceptable Deviance and Pearson statistics and a non-significant Hosmer–Lemeshow test. The probit model performs slightly worse: although its Deviance is similar, the significant Pearson and Hosmer–Lemeshow p-values indicate a lack of fit. The complementary log–log model provides the best overall fit, with favorable Deviance, a highly non-significant Pearson test, and a non-significant Hosmer–Lemeshow statistic, all suggesting a strong model fit. Therefore the null hypothesis that there is no significant difference in model performance among the three link functions is rejected. The complementary log-log was selected as the ideal appropriate theoretical model that best explained HIV determinant under this study as decided from the three goodness-of- fit techniques

Table 2: Parameter Estimate for Binary Logistic Regression Using Logit Functions

Term	Coefficient	Standard Error	Z-Value	P-Value	Odd Ratio exp(Coefficient)
Constant	-0.336	0.152	-2.20	0.028*	0.7146
YEAR 2018 (Ref)					1.000
2019	-1.444	0.133	-10.86	0.000*	0.2360
2020	-1.448	0.128	-11.27	0.000*	0.2350
2021	-1.060	0.131	-8.08	0.000*	0.3465
2022	-2.259	0.129	-17.55	0.000*	0.1045
2023	-1.189	0.141	-8.41	0.000*	0.3045
SEX Female (Ref)					1.000
Male	-0.6881	0.0664	-10.37	0.000*	0.5025
AGE 0-14 (Ref)					1.000
15-29 years	0.125	0.129	0.97	0.331	1.1331
30-49 years	2.240	0.122	18.34	0.000*	9.3933
50 years and above	1.881	0.143	13.17	0.000*	6.5601

* Significant at 5% Level

The parameter estimates presented in Table 3 indicate that the baseline odds of HIV infection, when all predictors are at their reference categories, are below 1 (0.7146), reflecting a reduced baseline likelihood of HIV. All years compared with the reference year (2018) show a significant decline in the odds of HIV, with the sharpest reduction observed in 2022 (odds ratio = 0.1045), suggesting a consistent decrease in HIV odds over time.

For gender, males have lower odds of HIV infection compared to females (odds ratio = 0.5025). In the age category 15–29 years, the odds ratio is 1.1331, though this effect is not statistically significant. However, the 30–49 age group shows a substantial and significant increase in odds (odds ratio = 9.3933), indicating a high likelihood of HIV infection. Similarly, individuals aged 50 years and above also exhibit elevated odds (6.5601), suggesting that older adults are more likely to contract HIV than those in the 0–14 age group.

Regression equation for logit

$$\log it\pi(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

=-0.336 + 0.0 year_1 - 1.444 year_2 - 1.448 year_3 - 1.060 year_4 - 2.259 year_5 - 1.189 year_6 + 0.0 sex_1 - 0.6881 sex_2 + 0.0 age_1 + 0.125 age_2 + 2.240 age_3 + 1.881 age_4

Table 3: Parameter Estimate for Binary Logistic Regression Using Probit Functions

Term	Coefficient	Standard Error	Z-Value	P-Value	Odd Ratio exp(Coefficient)
Constant	-0.1596	0.0856	-1.87	0.062	0.8525
YEAR 2018 (Ref)					1.000
2019	-0.8589	0.0776	-11.07	0.000*	0.4236
2020	-0.8625	0.0750	-11.51	0.000*	0.4221
2021	-0.6439	0.0769	-8.37	0.000*	0.5252
2022	-1.3115	0.0738	-17.77	0.000*	0.2694
2023	-0.6860	0.0821	-8.36	0.000*	0.5036
SEX Female (Ref)					1.000
Male	-0.4006	0.0385	-10.41	0.000*	0.6699
AGE 0-14 (Ref)					1.000
15-29 years	0.0399	0.0682	0.58	0.559	1.040
30-49 years	1.2842	0.0654	19.65	0.000*	3.6118
50 years and above	1.0586	0.0792	13.36	0.000*	2.8823

* Significant at 5% Level

Table 3 above presents parameter estimates for the probit model. The constant term shows a similar pattern to the logit model but is slightly less significant (p = 0.062). For the year variable, all years compared to the reference year (2018) show significantly reduced odds of HIV infection: 2019 (0.4236), 2020 (0.4221), 2021 (0.5252), 2022 (0.2694), and 2023 (0.5036). All odds ratios are significant at the 0.05 level, indicating a consistent decline across the years. These results align with the logit model, although the odds ratios are slightly smaller.

For gender, using females as the reference group, males have lower odds of HIV infection (0.6699). Regarding age, with the 0–14 group as the reference, the odds for ages 15–29 increase slightly (1.040) but are not statistically significant. The 30–49 age group shows a significant increase in odds (3.6118), while individuals aged 50 and above also have significantly higher odds (2.8823) of HIV infection.

Regression equation for probit model

$$\pi(x) = \phi(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$$

$\varphi^{-1}\pi(x)$ =-0.1596 + 0.0 year_1 - 0.8589 year_2 - 0.8625 year_3 - 0.6439 year_4 - 1.3115 year_5
- 0.6860 year_6 + 0.0 sex_1 - 0.4006 sex_2 + 0.0 age_1 + 0.0399 age_2 + 1.2842age3
+ 1.0586 age4

Φ= CDF of the standard normal distribution

Table 4: Parameter Estimate for Binary Logistic Regression Using Complementary Log-Log Functions

Term	Coefficient	Standard Error	Z-Value	P-Value	Odd Ratio exp(Coefficient)
Constant	-0.946	0.120	-7.90	0.000*	0.3883
YEAR 2018 (Ref)					1.000
2019	-0.9232	0.0845	-10.93	0.000*	0.3972
2020	-0.9251	0.0807	-11.46	0.000*	0.3965
2021	-0.6193	0.0826	-7.50	0.000*	0.5383
2022	-1.6019	0.0852	-18.80	0.000*	0.2015
2023	-0.8117	0.0877	-9.26	0.000*	0.4441
SEX Female (Ref)					1.000
Male	-0.4988	0.0498	-10.01	0.000*	0.6073
AGE 0-14 (Ref)					1.000
15-29 years	0.128	0.117	1.09	0.276	1.1365
30-49 years	1.825	0.107	16.99	0.000*	6.2028
50 years and above	1.568	0.121	13.01	0.000*	4.7970

* Significant at 5% Level

Complementary log–log model parameter estimates shown in Table 4 above reveal that, The constant term shows the baseline odds of HIV infection, when all predictors are at their reference levels, are below 1 (0.3883), indicating a reduced initial likelihood of HIV. All years compared to the reference year (2018) have odds ratios less than 1, demonstrating a consistent decrease in the odds of HIV infection over time.

For gender, males have lower odds of HIV compared to females (0.6073). Regarding age, individuals aged 15–29 have slightly higher odds of HIV (1.1365) relative to those aged 0–14, though this increase is not statistically significant. The 30–49 age groups shows a significant increase in odds (6.2028), and those aged 50 and above also exhibit higher odds (4.7970), indicating greater likelihood of HIV infection compared to the youngest age group.

Regression Equation For complementary Log log Model

$$\pi(x) = 1 - \exp(-\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3))$$

$\pi(x)$ =-0.946 + 0.0 year_1 - 0.9232 year_2 - 0.9251 year_3 - 0.6193 year_4 - 1.6019 year_5 - 0.8117 year_6 + 0.0 sex_1 - 0.4988 sex_2 + 0.0 age_1 + 0.128 age_2 + 1.825 age_3 + 1.568 age_4

Table 5: Model Selection for Binary Logistic Regression Using Various Link Functions

Link Function	AIC	BIC
Logit	6086.35	6153.79
Probit	6095.31	6162.75
Complementary Log-Log	6083.23*	6150.66*

* Best fit Model

Model Selection

In this study to know the best model fit between the various links functions (logit, probit and complementary log log or Gombit model). The Akaike information criterion and the Bayesian information criteria was use in selecting the optimal model. The optimal model is the model having a minimal value of the AIC and BIC. (Agresti, 2007)

In table 5 above the model with the best fit is the complementary log log because it has the smallest AIC and BIC value.

Conclusion

This study applied binary logistics regression using three link functions (logit. Probit, and complementary log-log) on a dataset of 6,268 observations, of which 1,995 (31.8%) were HIV positive obtained from Haruna Tsokwa memorial hospital (general hospital) Takum, Taraba State Nigeria. To assessed the significant determinants of HIV infection and as well identified the model that best fit the data. Goodness-of-fit results indicate that all models performed reasonably well, but the complementary log–log model showed the strongest fit, with non-significant goodness-of-fit tests and the lowest AIC and BIC values. Across all models, HIV odds declined significantly from 2019 to 2023 relative to 2018, suggesting improvements in prevention, awareness, and treatment efforts. Females consistently exhibited higher odds of HIV infection than males, reflecting gender-based vulnerabilities documented in previous studies. Age was a strong predictor, with individuals aged 30–49 and 50 years and above showing significantly higher odds of infection compared to those aged 0–14 years. The complementary log–log model outperformed the logit and probit models, likely due to its ability to capture asymmetry in the probability distribution of HIV infection. This finding supports the use of alternative link functions in epidemiological modeling, particularly when event probabilities are unevenly distributed.

This study demonstrates that age, sex, and year are significant determinants of HIV infection among patients in General Hospital Takum. The declining trend in HIV odds over time indicates progress in HIV control efforts, although females and adults aged 30 years and above remain at higher risk. Among the three models examined, the complementary log–log model provided the best fit and is recommended for predicting HIV infection in similar datasets.

Recommendations

Based on the finding of this study the following recommendation is made

1. Strengthening targeted interventions for high-risk age groups.
2. Enhancing gender-focused HIV prevention programs.
3. Sustaining investments in HIV testing and treatment.
4. Encouraging the use of alternative link functions in future epidemiological studies.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Wiley. https://doi.org/10.1002/0471249688

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Wiley. https://doi.org/10.1002/0470114754

Akinrefon, A. A., Emmanuel, R., & Okolo, A. (2023). Log-linear models for HIV/AIDS prevalence in Adamawa State, Nigeria. *FUDMA Journal of Sciences, 7*(3), 103–109. https://doi.org/10.33003/fjs-2023-0703-1804

Awofala, A. A., & Ogundele, O. E. (2018). HIV epidemiology in Nigeria. *Saudi Journal of Biological Sciences, 25*(4), 697–703. https://doi.org/10.1016/j.sjbs.2016.03.006

Awoleye, O. J., & Thron, C. (2015). Determinants of human immunodeficiency virus (HIV) infection in Nigeria: A synthesis of the literature. *Journal of AIDS and HIV Research, 7*(9), 117–129. https://doi.org/10.5897/JAHR2015.0338

Gűriş, S., Çağlayan, E., & Ün, T. (2011). Estimating of probability of home-ownership in rural and urban areas: Logit, probit, and gompit model. *European Journal of Social Sciences, 21*(3), 405–411.

Kamal, A., & Pervaiz, M. K. (2011). Factors affecting the family size in Pakistan: Clog-log regression model analysis. *Journal of Statistics, 18*(1), 29–53.

Kudakwashe, M., & Yesuf, K. M. (2014). Application of binary logistic regression in assessing risk factors affecting the prevalence of toxoplasmosis. *American Journal of Applied Mathematics and Statistics, 2*(6), 357–363. https://doi.org/10.12691/ajams-2-6-1

Obeagu, E. I., & Obeagu, G. U. (2022). An update on survival of people living with HIV in Nigeria. *Journal of Public Health and Nutrition, 5*(6), Article 129. https://doi.org/10.35841/aajphn-5.6.129

Ochalek, J., Revill, P., & van den Berg, B. (2017). Causal effects of HIV on employment status in low-income settings. *Economics & Human Biology, 27*(A), 248–260. https://doi.org/10.1016/j.ehb.2017.09.001

Odimegwu, C. O., Akinyemi, J. O., & Alabi, O. O. (2017). HIV-stigma in Nigeria: Review of research studies, policies, and programmes. *AIDS Research and Treatment, 2017*, Article 5812650. https://doi.org/10.1155/2017/5812650

Onemayin, K. J., Yusuf, H. O., Oluwafemi, S. O., & Abiodun, A. A. (2019). An assessment of HIV counselling and testing (HCT) service utilization in Nigeria: A binary logistic regression approach. *International Journal of HIV/AIDS Prevention, Education and Behavioural Science, 5*(1), 26–36. https://doi.org/10.11648/j.ijhpebs.20190501.14

Ramayani, R., Hamid, A., & Kismawadi, E. R. (2020). Pengaruh kepercayaan, keamanan, manfaat dan kemudahan terhadap penggunaan mobile banking. *JIM: Jurnal Ilmiah Mahasiswa, 2*(2), 93–108. https://doi.org/10.32505/jim.v2i2.2638

Ruspriyanty, D. I., & Sofro, A. (2018). Analysis of hypertension disease using logistic and probit regression. *Journal of Physics: Conference Series, 1108*, Article 012054. https://doi.org/10.1088/1742-6596/1108/1/012054

Seyoum, S. (2018). Analysis of prevalence of malaria and anemia using bivariate probit model. *Annals of Data Science, 5*(2), 301–312. https://doi.org/10.1007/s40745-018-0138-3

UNAIDS. (2024). *Global HIV & AIDS statistics — Fact sheet.* https://www.unaids.org/en/resources/fact-sheet