

Original Article

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

^{1*}Tugga, Hussaina. Ahmad., ²Ogunmola, Adeniyi. Oyewole.,
³Bamigbala, Olateju. Alao., ⁴Ahmad, Salim.Salim.

^{1,2,3}Department of Statistics, Federal University Wukari, Nigeria.

⁴Federal University Dutsin-Ma, Katsina State, Nigeria.

hussainatugga@fuwukari.edu.ng

Received: 15 December 2025; Revised: 16 January 2026; Accepted: 21 February 2026; Published: 08 March 2026

Abstract - Human Immunodeficiency Virus (HIV) remains one of the major public health concerns globally, with sub-Saharan Africa accounting for a significant proportion of the global burden of the infection (Kharsany & Karim, 2016). According to UNAIDS (2024), millions of people continue to live with HIV, with new infections occurring daily due to persistent demographic and behavioral factors. In Nigeria, the HIV epidemic exhibits geographic and demographic variation, influenced by age, sex, socioeconomic status, risk behaviors, and accessibility to healthcare services (Obeagu & Obeagu, 2022). Understanding the determinants of HIV infection is crucial for effective prevention, early detection, and policy formation. Statistical modeling plays a vital role in identifying and quantifying the effect of such determinants. This study applies binary logistic regression using logit, probit, and complementary log-log (cloglog) link functions to assess the influence of age, sex, and year on HIV infection among patients tested at General Hospital Takum, Taraba State, Nigeria, between 2018 and 2023. The objective is to identify significant demographic determinants of HIV infection and determine the best-fitting link function for the data. Model performance was evaluated using goodness-of-fit statistics (Deviance, Pearson chi-square, and Hosmer-Lemeshow tests) and model selection criteria (Akaike Information Criterion and Bayesian Information Criterion). Results indicate a consistent decline in HIV odds across the years, significantly higher odds among females, and substantially increased odds among adults aged 30–49 and 50 years and above. Among the three models, the complementary log-log link function demonstrated the best overall fit, exhibiting 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. The findings are expected to contribute to improved model selection in epidemiological research and inform localized HIV prevention strategies.

Keywords - HIV, Binary Logistic Regression, Logit, Probit, Complementary Log-Log, Model Selection, Epidemiology.

1. 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 (Kharsany & Karim, 2016). 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, inadequate access to healthcare services, and HIV-related stigma continue to undermine effective prevention and treatment strategies, while key populations, including female sex workers, injecting drug users, and men who have sex with men, although representing a small proportion of the population, contribute disproportionately to new HIV infections (Awolaye & Thron, 2015). 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 logistic regression is widely used for modelling 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, such as probit and complementary log-log, can also be used. Especially in situations 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.

2. 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.

3. Materials and Methods

3.1. 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.

3.2. 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).

3.3. Methodology

In this section, we discuss the binary regression model with various link functions (logit, probit, and the complementary log-log link functions), with each of these models using a different functional form to relate the probability p to the regressors, and they are all estimated using maximum likelihood methods (Guris & Caglyan, 2011). Binary regression models form the basis for more advanced models used for ordinal, nominal, and count data; the model analyzes how explanatory variables influence the probability of the event occurring (Kamal and Pervaiz, 2011).

3.4. Binary Logistic Regression Model

Logistic regression is a nonlinear model and a special case of a generalized linear model (GLM) with a binomial random component and a logit link function. It assumes a linear relationship between the explanatory variables and the log-odds of the response, producing predicted probabilities that range between 0 and 1 (Agresti, 2007). For a binary response variable Y and explanatory variable X , let

$$\pi(x) = P(Y = 1/X = x) \text{ (Agresti, 2007).}$$

The binary logistic regression model with multiple explanatory variables is expressed as follows:

For $\pi(x) = P(y = 1)$ with $X = (x_1, \dots, x_p)$

$$= \log \left[\frac{\pi(x)}{1-\pi(x)} \right] = \log \text{it}[\pi(x)] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \tag{1}$$

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

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

X_i 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. (Agresti, 2007).

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

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

Conversely, the probability of failure is,

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

3.5. Probit Model

Another nonlinear model used to analyze the relationship between a binary response variable and multiple explanatory variables is the probit model. It explains a dichotomous dependent variable using the Cumulative Distribution Function (CDF) of the normal distribution, and is therefore also referred to as the normit model (Ruspriyanty & Sofro, 2018). Let $\phi(\cdot)$ denote the standard CDF of a Normal distribution, such as the $N(0, 1)$ CDF, and the probit link function is $\Phi^{-1}(\cdot)$. The probit model is expressed as:

Probit

$$\pi(x) = \Phi^{-1}[\pi(x)] \tag{4}$$

3.6. Complementary Log-Log Model

The complementary log-log regression model is another binary regression model. The link function used in this Generalized Linear Model (GLM) is the complementary log-log link, which applies the log-log transformation to the complement of the probability (Kamal & Pervaiz, 2011). Complementary log-log Regression Model is expressed as;

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

Taking the log of the negative log of the complement of $\pi(x) = p(y = 1/X = x)$, the link function of the model can be expressed as follows:

$$\log\{-\log(1 - p)\} = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p \quad (6)$$

3.7. Inference for Logistic Regression

The likelihood ratio test is used 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 no 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 a significant difference in model performance among the three link functions.

3.8. Likelihood Ratio Test

To detect lack of fit in a binary response model, the likelihood ratio statistic is commonly used to test whether certain model parameters are zero by comparing the log-likelihood L_1 for the fitted model M_1 with L_0 that of a simpler model M_0 . Denote this statistic for testing M_0 , given M_1 that $G^2(M_0/M_1)$ 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 (Agresti, 2002). The likelihood ratio statistic for comparing M_1 and M_0 is

$$G^2(M_0/M_1) = G^2(M_0) - G^2(M_1) \quad (7)$$

3.9. Goodness of Fit Test

The Hosmer-Lemeshow test is a chi-square (χ^2) Goodness-of-fit test used to evaluate the adequacy of a model in fitting observed data (Qais M. Abdulqader, 2017). In addition, the quality of the fitted models can be assessed using Pearson's chi-square statistic and the likelihood ratio statistic (G^2), both of which compare observed counts with those predicted by the model. A significant chi-square value indicates that the variable contributes significantly to the model (Kudakwashe & Yesuf, 2014).

3.10. Binary Logistic Regression Assumption

1. The response variable can only take two possible outcomes.
2. There should be no multi-collinearity between the explanatory variables.
3. There should not be extreme outliers.
4. The log odds assume a linear relationship with the explanatory variable.
5. The response variables are mutually independent.

3.11. Linear Approximation Interpretation for β_i

The S-shaped form of $\pi(x)$ is determined by the rate at which the parameter β increases or decreases in equations (1) and (2). The sign of β indicates the direction of the curve: when $\beta > 0$, the curve is increasing as x increases, whereas when $\beta < 0$, the curve decreases as x increases. Furthermore, the rate of change becomes steeper as $|\beta|$ increases. When $\beta = 0$, the right-hand side of equation (1) reduces to a constant, implying that $\pi(x)$ remains the same for all values of x . In this case, the curve becomes a horizontal straight line, indicating that the binary response variable Y is independent of X (Agresti, 2007).

3.12. The Odds Ratio Interpretation

Equation (8) provides an interpretation of the odds of response in the model. Specifically, the odds of the response are multiplied by e^β for every one-unit increase in x . In other words, the odds at level $x + 1$ are equal to the odds at level x multiplied by e^β . When $\beta = 0$, $e^\beta = 1$, indicating that the odds remain unchanged regardless of the value of x (Agresti, 2007). Taking the exponential of both sides of equation (1), the odds of response are:

$$\frac{\pi(x)}{1-\pi(x)} = \exp(\alpha + \beta x) = e^\alpha (e^\beta)^x \tag{8}$$

4. Results and Discussion

4.1. Goodness of Fit Tests

To identify the appropriate link functions that fit the response variables, three goodness-of-fit techniques were adapted to assess the adequacy of these models on the data presented in 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 show 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 the 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 in Table 2 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 infection. 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 the logit function.

$$\text{logit } \pi(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$= -0.336 + 0.0 \text{ year}_1 - 1.444 \text{ year}_2 - 1.448 \text{ year}_3 - 1.060 \text{ year}_4 - 2.259 \text{ year}_5 - 1.189 \text{ year}_6 + 0.0 \text{ sex}_1 - 0.6881 \text{ sex}_2 + 0.0 \text{ age}_1 + 0.125 \text{ age}_2 + 2.240 \text{ age}_3 + 1.881 \text{ age}_4$$

Table 3. Parameter estimate for binary logistic regression using probit functions

Term	Coefficient	Standard Error	Z-Value	P-Value
Constant	-0.1596	s0.0856	-1.87	0.062
YEAR 2018 (Ref)				
2019	-0.8589	0.0776	-11.07	0.000*
2020	-0.8625	0.0750	-11.51	0.000*
2021	-0.6439	0.0769	-8.37	0.000*
2022	-1.3115	0.0738	-17.77	0.000*
2023	-0.6860	0.0821	-8.36	0.000*
SEX Female (Ref)				
Male	-0.4006	0.0385	-10.41	0.000*
AGE 0-14 (Ref)				

15-29 years	0.0399	0.0682	0.58	0.559
30-49 years	1.2842	0.0654	19.65	0.000*
50 years and above	1.0586	0.0792	13.36	0.000*

*Significant at 5% Level

Table 3 above presents parameter estimates for the probit model. The intercept (constant term) is negative (-0.1596) and marginally insignificant at the 5% level ($p = 0.062$), indicating that when all predictors are at their reference categories (year = 2018, female, age 0–14), the baseline latent propensity for HIV infection is slightly below zero, though not statistically distinguishable from zero at the conventional level. For the year variable, all categories (2019–2023) have negative and highly significant coefficients ($p < 0.001$) relative to the reference year (2018). This suggests a statistically significant decline in the probability of HIV infection in each of these years compared to 2018. Among them, 2022 shows the largest negative effect (-1.3115), indicating the strongest reduction in the latent propensity for HIV infection during that year. Regarding sex, the coefficient for males (-0.4006) is negative and statistically significant ($p < 0.001$), implying that being male is associated with a lower probability of HIV infection compared to females, holding other variables constant. For the age variable, individuals aged 15–29 years have a small positive coefficient (0.0399) that is not statistically significant ($p = 0.559$), indicating no meaningful difference in HIV risk compared to the reference group (0–14 years). However, individuals aged 30–49 years (1.2842) and 50 years and above (1.0586) have large, positive, and statistically significant coefficients ($p < 0.001$), suggesting that these age groups have a substantially higher probability of HIV infection relative to children aged 0–14 years. Regression equation for the probit model,

$$\pi(x) = \phi(\alpha + \beta_1x_1 + \beta_2x_2 + \beta_3x_3)$$

$$\phi^{-1}\pi(x) = -0.1596 + 0.0 \text{ year}_1 - 0.8589 \text{ year}_2 - 0.8625 \text{ year}_3 - 0.6439 \text{ year}_4 - 1.3115 \text{ year}_5 - 0.6860 \text{ year}_6 + 0.0 \text{ sex}_1 - 0.4006 \text{ sex}_2 + 0.0 \text{ age}_1 + 0.0399 \text{ age}_2 + 1.2842 \text{ age}_3 + 1.0586 \text{ age}_4$$

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 indicates 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 group shows a significant increase in odds (6.2028), and those aged 50 and above also exhibit higher odds (4.7970), indicating a greater likelihood of HIV infection compared to the youngest age group.

4.2. Regression Equation for Complementary Log-Log Model

$$\pi(x) = 1 - \exp(-\exp(\alpha + \beta_1x_1 + \beta_2x_2 + \beta_3x_3))$$

$$\pi(x) = -0.946 + 0.0 \text{ year}_1 - 0.9232 \text{ year}_2 - 0.9251 \text{ year}_3 - 0.6193 \text{ year}_4 - 1.6019 \text{ year}_5 - 0.8117 \text{ year}_6 + 0.0 \text{ sex}_1 - 0.4988 \text{ sex}_2 + 0.0 \text{ age}_1 + 0.128 \text{ age}_2 + 1.825 \text{ age}_3 + 1.568 \text{ age}_4$$

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

4.3. Model Selection

In this study, the authors aim to determine the best model fit between the various link functions (logit, probit, and complementary log–log or Gombit model). The Akaike information criterion and the Bayesian information criterion were used 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.

5. Conclusion

This study applied binary logistic regression using three link functions (logit, Probit, and complementary log–log) to 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 assess the significant determinants of HIV infection and to identify the model that best fits 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 modelling, 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.

5.1. Recommendations

Based on the findings 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

- [1] Awoyemi Abayomi Awofala, and Olusegun Emmanuel Ogundele, "HIV Epidemiology in Nigeria," *Saudi Journal of Biological Sciences*, vol. 25, no. 4, pp. 697-703, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Olatunji Awoloye, and Chris Thron, "Determinant of HIV in Nigeria, A Synthesis of the Literature," *Journal of AIDS and HIV Research*, vol. 7, no. 9, pp. 117-129, 2015. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Alan Agresti, *Categorical Data Analysis*, 3rd ed., John Wailey and Sons, Inc., Publication, 2013. [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Alan Agresti, *Introduction to Categorical Data Analysis*, John Wailey and Sons, Inc., Publication, 2006. [Online]. Available: <https://onlinelibrary.wiley.com/doi/book/10.1002/0470114754>
- [5] D.I. Ruspriyanty, and A. Sofro, "Analysis of Hypertension Disease using Logistic and Probit Regression," *Journal of Physics: Conference Series*, vol. 1108, no. 1, pp. 1-6, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Adesupo Adeoye Akinrefon, Rimamkyaten Emmanuel, and Abraham Okolo, "Log-Linear Models for HIV/AIDS Prevalence in Adamawa State, Nigeria," *FUDMA Journal of Sciences (FJS)*, vol. 7, no. 3, pp. 103-109, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Selahattin Güriş, Ebru Çağlayan, and Turgut Ün, "Estimating the Probability of Home-Ownership in Rural and Urban Areas: Logit, Probit, and Gompit Model," *European Journal of Social Sciences*, vol. 21, no. 3, pp. 405-411, 2011. [[Google Scholar](#)]
- [8] Asifa Kamal, and Muhammad Khalid Pervaiz, "Factors Affecting the Family Size in Pakistan: Clog-Log Regression Model Analysis," *Journal of Statistics*, vol. 18, 2011. [[Google Scholar](#)]
- [9] Mutangi Kudakwashe, and Kasim Mohammed Yesuf, "Application of Binary Logistic Regression in Assessing Risk Factors Affecting the Prevalence of Toxoplasmosis," *American Journal of Applied Mathematics and Statistics*, vol. 2, no. 6, pp. 357-363, 2014. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Ayesha B.M. Kharsany, and Quarraisha A. Karim, "HIV Infection and AIDS in Sub-Saharan Africa: Current Status, Challenges and Opportunities," *The Open AIDS Journal*, vol. 10, pp. 34-48, 2016. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Clifford O. Odimegwu, Joshua O. Akinyemi, and Olatunji O. Alabi, "HIV-Stigma in Nigeria: Review of Research Studies, Policies, and Programmes," *AIDS Research and Treatment*, vol. 2017, no. 1, pp. 1-12, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Emmanuel Ifeanyi Obeagu, and Getrude Uzoma Obeagu, "An Update on Survival of People Living with HIV in Nigeria," *Journal of Public Health Nutrition*, vol. 5, no. 6, pp. 1-8, 2022. [[CrossRef](#)] [[Google Scholar](#)]
- [13] Onemayin Kolawole Josph et al., "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*, vol. 5, no. 1, pp. 26-36, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Jessica Ochalek, Paul Revill, and Bernard Van Den Berg, "Causal Effects of HIV on Employment Status in Low-Income Settings," *Economics and Human Biology*, vol. 27, pp. 248-260, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Senayit Seyoum, "Analysis of Prevalence of Malaria and Anemia using Bivariate Probit Model," *Annals of Data Science*, vol. 5, no. 2, pp. 301-312, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] UNAIDS, *Global HIV and AIDS Statistics*, 2024. [Online]. Available: <https://www.unaids.org/en/resources/fact-sheet>