American Journal of Theoretical and Applied Statistics
2024; 13(4): 80-84
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20241304.13
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

A Generalized Linear Model of HIV/AIDS Patients in Kenya: A Case Study of Nyeri County Referral Hospital

Sarah Ireri[1, *], Joseph Esekon[1], Margaret Kinyua[2]

[1]Department of Pure and Applied Sciences, Kirinyaga University, Kerugoya, Kenya
[2]Department of Mathematics, Statistics and Actuarial Science, Karatina University, Karatina, Kenya

Email address:

smukami1001@gmail.com (Sarah Ireri), jesekon@kyu.ac.ke (Joseph Esekon), mwambui@karu.ac.ke (Margaret Kinyua)
[*]Corresponding author

To cite this article:

Sarah Ireri, Joseph Esekon, Margaret Kinyua. (2024). A Generalized Linear Model of HIV/AIDS Patients in Kenya: A Case Study of Nyeri County Referral Hospital. *American Journal of Theoretical and Applied Statistics, 13*(4), 80-84. https://doi.org/10.11648/j.ajtas.20241304.13

Received: 24 July 2024; **Accepted:** 13 August 2024; **Published:** 22 August 2024

Abstract: With millions of new cases and deaths reported every year, HIV/AIDS is a significant worldwide health concern. Creating successful public health policies and interventions requires an understanding of the dynamics of HIV transmission and progression. WHO predicted that by the end of 2022 roughly 39 million individuals worldwide would be living with HIV, out of which 37.5 million are adults, whereas 1.5 million are children. Despite outstanding global gains in HIV/AIDS prevention, treatment and care, Kenya continues to struggle to effectively handle the HIV epidemic, particularly in areas like Nyeri County. Nyeri County Referral Hospital is a critical healthcare institution for HIV/AIDS patients in the region. However, there is still a lack of understanding about the epidemiological characteristics of HIV/AIDS in this particular population. This study's aim was to use a GLM on HIV/AIDS data in Nyeri County Referral Hospital in Kenya. To determine the significance of model parameters, Likelihood Ratio Test was used whereas significance of regression coefficients was determined using Wald Chi-Square Test. Deviance was utilized to test for the goodness of fit. R software version 4.4.1 was utilized. This project may help health policymakers in developing or refining HIV/AIDS care programs. Findings from the study can help healthcare planners and policymakers allocate resources more efficiently to meet the requirements of HIV/AIDS patients. The fitted model showed that, only ART use was significant (p-value = 2.684562×10^{-13}). Because some covariates were not significant, each of them was analyzed separately. Age was a significant predictor (p-value = 0.0001536103). The other variables were not significant. This finding is consistent with previous evidence, which stresses the relevance of ART in lowering viral load, enhancing immunological function, and extending the lives of people living with HIV. To build upon the current findings, future research should explore additional variables that may influence HIV status, for example cultural beliefs, and access to healthcare services. Again, future studies may involve the use of survival analysis through GLM in analyzing similar data.

Keywords: Generalized Linear Model, HIV/AIDS, Logistic Regression, Likelihood Ratio Test

1. Introduction

The HIV/AIDS is among the major global health problems with every year millions of people coming under its influence while millions are dying from it around the face of the world. Health providers will not be able to develop effective public health policies as well as programs and interventions without understanding the HIV transmission and progression factors [2]. For many people, HIV has become less severe since the discovery of antiretroviral therapy (ART), turning it from a fatal sickness to a chronic condition that can be treated [1]. We utilize a flexible model, in this context a generalized linear model (GLM), which provides a means for the incorporation of a multitude of the features which cause or accelerate the HIV transmission and get an individual in conviction with the disease. In countries with low resources, the number of HIV patients undergoing treatment has significantly increased during the past ten years. By the end of December 2022, 29.8 million persons living with HIV globally had access to ART, up from 7.7 million in 2010 [15]. Despite these remarkable accomplishments, HIV remains the world's greatest infectious killer.

It was the goal of this research project to investigate the mechanics of HIV infection and bringing its evidence-based interventions to light through diversity concerning the use of GLMs in the epidemiology. The struggle with HIV has become the world priority, and its goal is to search for the cure, prevention, and treatment after being afflicted. Especially in low- and middle-income nations, the illness is among the most pressing issues in public health in the globe [7].

In many regions of the world, poverty, illiteracy and restricted access to healthcare services pose serious obstacles to HIV prevention and treatment. Discrimination and stigma against those who are HIV/AIDS positive might deter people from getting tested and treated, which accelerates the virus's transmission [2]. HIV transmission is largely caused by behavioral factors, including multiple sexual partners, unprotected intercourse, and intravenous drug use. Developing focused strategies to lower the incidence of HIV requires an understanding of these factors [4].

There is still a lack of understanding about the epidemiological characteristics of HIV/AIDS [14]. Individuals who are HIV positive are more vulnerable to secondary infections because of their compromised immune systems. One of the most prevalent co-infections and a major cause of death for HIV-positive individuals is tuberculosis (TB). HIV-positive people are also frequently infected with human papillomavirus (HPV), hepatitis B and C and other sexually transmitted infections (STIs). These co-infections make treatment and management more difficult, necessitating the use of integrated healthcare approaches to treat several illnesses at once [5, 8].

Applying the GLM to the data, we obtained the results in Table 1. ART use had a significant positive effect on the outcome (p-value = 2.684562×10^{-13}). This means that, holding all other factors constant, individuals using ART are much more likely to have a positive HIV status compared to those not using ART. The negative estimate suggested that as age increases, the likelihood of the outcome decreases. In the multivariable model, ART use remained a highly significant predictor, showing a strong positive effect on the outcome. Age, however, was not significant when it is included in the model containing ART use.

2. Materials and Methods

Generalized Linear Model

Generalized Linear Models are essential in broadening the understanding of linear regression models. GLMs also expand the general linear model framework to meet instances when general linear models are inapplicable[6,11]. Generalized Linear Models (GLMs) come with several assumptions. Firstly, the data points $(Y_1, Y_2, \cdots Y_n)$ are assumed to be independently distributed. Secondly, while GLMs often use distributions from the exponential family, the dependent variable Y_i doesn't necessarily need to be normally distributed [9,10]. In this work, we considered the Logistic Regression model for modelling HIV Status. The model characteristics included [3], a linear predictor given by

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = \beta_0 + \sum_{j=1}^{p} \beta_i x_{ij}, \quad i = 1, \cdots, n \tag{1}$$

and a Link function which explains the correlation between the mean, $E(Y_i) = \mu_i$, and the linear predictor, $g(\mu_i) = \eta_i$.

The logit model is given as

$$P(Y = 1 \mid X = x_i) = \frac{\exp(\beta_0 + \sum_{i=1}^{p} \beta_i x_i)}{1 + \exp(\beta_0 + \sum_{i=1}^{p} \beta_i x_i)}. \tag{2}$$

The logit transformation allowed modeling of the binary outcome as a linear combination of predictor variables. By transforming the probabilities to the log-odds scale, it linearized the relationship between the predictors and the HIV Status. That is

$$\text{logit}(P(x)) = \ln\left[\frac{P(Y = 1 \mid X = x_i)}{1 - P(Y = 1 \mid X = x_i)}\right] \tag{3}$$

When Y contains binary data, the correlation between the numerical parameters and the response variable $Y = f(x)$ can be fitted using a logistic regression curve. The logistic curve first grows gradually before becoming exponential and then settling at a consistent pace [12].

3. Results and Discussion

3.1. Parameter Estimates of the Logistic Regression Model

The generalized linear model was fitted to determine which variables are significant in modelling the HIV/AIDS status (Table 1).

Table 1. *Parameter Estimates from the Logistic Regression Model.*

Coefficient	Estimate	Standard Error	z value	$\Pr(>\|z\|)$
Intercept	-14.11691577	2.95389118	-4.7790913	1.760892×10^{-6}
Age	-0.02540454	0.01850866	-1.3725761	1.698842×10^{-1}
Gender	0.23198580	0.51200343	0.4530942	6.504809×10^{-1}
Marital Status	0.14431027	0.51823771	0.2784635	7.806566×10^{-1}
Education level	0.12389535	0.25770584	0.4807627	6.306852×10^{-1}
Alcohol use	0.15268775	0.53409249	0.2858826	7.749680×10^{-1}
Co-infection	0.19989244	0.14026956	1.4250593	1.541401×10^{-1}
ART use	7.70583622	1.05424504	7.3093407	2.684562×10^{-13}

The intercept is highly significant with a very small p-value (1.760892×10^{-6}), indicating that the baseline log-odds of the response variable (when all predictors are zero) is significantly different from zero. ART use has a significant positive effect on the outcome (p-value = 2.684562×10^{-13}). This means that, holding all other factors constant, individuals using ART are much more likely to have a positive HIV status compared to those not using ART. This is an important finding as this indicates that, the administration of the ART on the HIV positive individuals is actually on the targeted population.

Some covariates were not significant. An analysis for each of them separately apart from Model 8 which had two covariates was conducted, (See Table 2).

3.2. Logistic Regression Results for Different Models

The results for the logistic Regression for the different Models are displayed in Table 2.

Table 2. *Logistic Regression Results for Different Models.*

Coefficient	Estimate	Std. Error	z-value	$\Pr(>\|z\|)$
Model 1				
Intercept	0.54534063	0.333430530	1.635545	0.1019348537
Age	-0.02969899	0.007846168	-3.785159	0.0001536103
Model 2				
Intercept	-0.63991659	0.3355354	-1.9071510	0.05650104
Gender	-0.02292524	0.2114969	-0.1083951	0.91368225
Model 3				
Intercept	-0.59288690	0.3376967	-1.7556787	0.07914328
Marital Status	-0.05374026	0.2116011	-0.2539697	0.79951894
Model 4				
Intercept	-1.0315227	0.3717983	-2.774415	0.005530106
Alcohol use	0.2205925	0.2189891	1.007322	0.313780179
Model 5				
Intercept	-0.97024046	0.2386244	-4.065974	4.783234×10^{-5}
Co-infection	0.08121534	0.0579919	1.400460	1.613756×10^{-1}
Model 6				
Intercept	-13.091484	2.020490	-6.479362	9.211110×10^{-11}
ART use	7.578055	1.034557	7.324924	2.390338×10^{-13}
Model 7				
Intercept	-0.56932	0.31745	-1.793	0.0729
Education level	-0.03961	0.11300	-0.351	0.7260
Model 8				
Intercept	-11.92245496	2.11544245	-5.635916	1.741307×10^{-8}
Age	-0.02816332	0.01831381	-1.537819	1.240929×10^{-1}
ART use	7.56994389	1.03860089	7.288598	3.131981×10^{-13}

In the univariable Model 1, Age is a significant predictor (p-value = 0.0001536103). The negative estimate suggests that as age increases, the likelihood of the outcome decreases. This could be due to an individual becoming responsible as one advances in age. Gender, Marital status, Education level, Alcohol use and Co-infection are not significant predictors when considered alone. However, ART use was a highly significant predictor in this univariable model (p-value = 2.390338×10^{-13}). The large positive estimate indicates a strong effect on the outcome, with ART use greatly increasing the likelihood of the ART being used for the HIV positive individuals as intended. In this multivariable model, ART use remained a highly significant predictor, showing a strong positive effect on the outcome. Age, however, is not significant when it is included in the model containing ART use (See Model 8). This suggests that while age might be significant in a univariable analysis, its effect might be overshadowed by the strong influence of ART use in the multivariable context.

4. Conclusions

The study found that ART use was the most significant predictor of HIV status in the sample population (p-value = 2.684562×10^{-13}). Individuals on ART were likely to be HIV positive, highlighting the critical importance of ART in regulating HIV transmission and progression. When ART use was taken into account, however, age lost its statistical significance, implying that its impact on HIV status is moderated in the face of successful treatment. This finding suggests that, regardless of age, access to and use of ART is critical in HIV management. It also emphasizes the potential of ART to improve HIV-related health outcomes across age groups. Other variables such as gender, marital status, education level, alcohol use, and co-infection with diseases like tuberculosis were not individually significant in the multivariable model. In the univariable Model 1, Age was a significant predictor (p-value = 0.0001536103).

In conclusion, to build upon the current findings, future research should explore additional variables such as cultural beliefs, and access to healthcare services that may influence HIV status. Longitudinal studies may also be considered in future studies.

Acknowledgments

The main author would like to express her gratitude to [Dr. Joseph Eyang'an Esekon & Dr. Margaret Wambui Kinyua)] for their invaluable technical assistance and support throughout the research process. Special thanks are also extended to [Kirinyaga University] for providing essential resources and logistical aid. The contributions of these individuals and organizations were crucial to the successful completion of this research.

Author Contributions

Sarah Ireri: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing

Joseph Esekon: Supervision

Margaret Kinyua: Conceptualization, Formal Analysis, Methodology, Supervision, Writing - review & editing

ORCID

0009-00002-8288-4438 (Sarah Ireri)
0009-0000-8033-9175 (Joseph Esekon)
0009-0000-8317-2147 (Margaret Kinyua)

Funding

This study was self-funded. The author(s) received no particular grants from public, commercial, or non-profit funding bodies to support the study or cover publication costs.

Conflicts of Interest

The authors declare no conflicts of interest. There are no financial, commercial, or other affiliations that could be perceived as potential conflicts of interest by the academic community.

References

[1] Deeks, S. G., Lewin, S. R., & Havlir, D. V. (2013). The end of AIDS: HIV infection as a chronic disease. The lancet, 382(9903), 1525-1533.

[2] Granich, R., Gupta, S., Hersh, B., Williams, B., Montaner, J., Young, B., & Zuniga, J. M. (2015). Trends in AIDS deaths, new infections and ART coverage in the top 30 countries with the highest AIDS mortality burden; 1990-2013. PloS one, 10(7), e0131353.

[3] Hastie, T. J., & Pregibon, D. (2017). Generalized linear models. In Statistical models in S (pp. 195-247). Routledge.

[4] Hoosen, N. (2021). Interventions for Improving Adherence and Retention in HIV-Infected Women on ART During Antenatal and Postnatal Care: A Systematic Review.

[5] Keats, E. C., Macharia, W., Singh, N. S., Akseer, N., Ravishankar, N., Ngugi, A. K., ... & Bhutta, Z. A. (2018). Accelerating Kenya's progress to 2030: understanding the determinants of under-five mortality from 1990 to 2015. BMJ global health, 3(3), e000655.

[6] Kearns, B., Stevenson, M. D., Triantafyllopoulos, K., & Manca, A. (2019). Generalized linear models for flexible parametric modeling of the hazard function. Medical Decision Making, 39(7), 867-878.

[7] Kiweewa, F., Esber, A., Musingye, E., Reed, D., Crowell, T. A., Cham, F., ... & Kibuuka, H. (2019). HIV virologic failure and its predictors among HIV-infected adults on antiretroviral therapy in the African Cohort Study. PloS one, 14(2), e0211344.

[8] Klatt, N. R., Chomont, N., Douek, D. C., & Deeks, S. G. (2013). Immune activation and HIV persistence: implications for curative approaches to HIV infection. Immunological reviews, 254(1), 326-342.

[9] McCullagh, P. (2019). Generalized linear models. Routledge.

[10] McCue, T., Carruthers, E., Dawe, J., Liu, S., Robar, A., & Johnson, K. (2008). Evaluation of generalized linear model assumptions using randomization. Unpublished manuscript. Retrieved from http://www.mun.ca/biology/dschneider/b7932/B7932 Final10Dec2008.pdf

[11] Neuhaus, J., & McCulloch, C. (2011). Generalized linear models. Wiley Interdisciplinary Reviews: Computational Statistics, 3(5), 407-413.

[12] Strickland, J. (2017). Logistic regression inside and out. Lulu.com

[13] Vicente, V., & Aguiar, P. (2023). Practical Epidemiology with Generalized Linear Models. Leya.

[14] Wanjiru, s. w. (2021). Efficacy of strategies that mitigate challenges faced by women infected with HIV/AIDS in Majengo urban informal settlement, Nyeri county, Kenya (doctoral dissertation, school of humanities and social sciences in partial fulfillment of the requirement for the degree of (Master of Arts) in (gender and development studies), Kenyatta university).

[15] World Health Organization. (2022). Health at a Glance: Asia/Pacific 2022 Measuring Progress Towards Universal Health Coverage: Measuring Progress Towards Universal Health Coverage. OECD publishing.