Vol. 4 (Iss. 2) 2023, pp. 984-992 African Journal of Empirical Research https://ajernet.net ISSN 2709-2607

# Comparison of the Performance of Logistic Regression Model in the Presence and Absence of Mediation

Ruth Naomi Wanga[1]
Dr David Anekeya Alilah[2]
Dr Everlyne Akoth Odero[3]

[1]wangaruth24@gmail.com
[2]dalila@mmust.ac.ke
[3]eodero@mmust.ac.ke

[1,2,3]Department of Mathematics Masinde Muliro University of Science and Technology, P.O Box 190-50100, Kakamega, Kenya

**ABSTRACT**

*Over the last decade major global efforts mounted to address the HIV epidemic has realized notable successes in combating the pandemic. Sub Saharan Africa (SSA) still remains a global epicenter of the disease, accounting for more than 70% of the global burden of infections. Despite wide spread use of various intervention strategies that act as mediation factors in Human Immunodeficiency Virus (HIV) prevention, HIV prevalence still remains a challenge especially in some geographic areas and populations. Therefore, how mediation factors interact with the prevailing HIV risk factors to cause an impact on its prevalence remains a question not answered. This study considered Exposure to HIV related media as a mediator variable in the relationship between HIV risk factors and HIV prevalence. Two logistic regression models, one in presence of mediation and another in absence of mediation were formulated and compared to establish the best performing model. Models were fitted to real data from the Kenya Population-based HIV Impact Assessment survey-2018 and model parameters were estimated using Maximum Likelihood Estimation in R. Results based on both Akaike's Information Criterion and the McFadden's $R^2$ value revealed that the model formulated in presence of mediation performed better compared to that without mediation.*

**Keywords:** HIV Prevalence, Logistic Regression, Mediation

## I. INTRODUCTION

According to the UNAIDS Global Report (2021), 39.0 million people worldwide were living with HIV/AIDS, with 1.3 million being new infections. Over two-thirds of all these people living with HIV/AIDS are found in sub-Saharan Africa. In 2017, Kenya was ranked as one of the countries hardest hit by the HIV/AIDS epidemic in terms of the estimated number of new HIV infections among adults aged 15 years and older; however, in recent years, Kenya has recorded steady progress in HIV/AIDS prevention (UNAIDS, 2017).

Prevention of the HIV epidemic, like other infectious diseases, depends on having a good understanding of the determinants of the spread of the infections and being able to explain their trends in disease magnitude and the evaluation of intervention programs (DHS Working Papers, 2017).

There has been a steady decline in HIV prevalence in Kenya due to various intervention strategies applied by both the national government and donor funding (Higa et al., 2022). However, there still remain high rates of new infections and differences in the risk of infection, due to the varying effects of interventions used on HIV/AIDS control, either directly or indirectly (Musyoki, 2017).

According to LaCroix et al. (2014), mass media interventions are useful in reducing global HIV/AIDS disparities because of their wide reach, standardization and repetition of messages, and the ability to use different content formats, including entertainment, news, and short advertisements or announcements. Intervention strategies such as exposure to HIV-related mass media are some of the underlying factors believed to mediate the relationship between HIV risk factors and HIV prevalence.

Many studies, such as LaCroix et al. (2014), Agha (2003), and Mugoya (2016), have addressed the relationship between the mediator variable and HIV prevalence, but little is known on how different risk factors interact with the

mediator to reduce HIV prevalence, considering that exposure to HIV-related mass media is a mediator variable between risk factors and HIV prevalence.

A path diagram, as indicated in Fig. 1 by Namazi (2016), describes a simple relationship between the dependent Y, mediator M, and independent X variables.

Namazi (2016), concludes that one route connects *X* and *Y* directly and is known as the direct effect of *X* to *Y*, whereas the other route connects *X* and *Y* via a mediator *M* and is known as the indirect effect of *X* to *Y*.

$\alpha$ is the total effect while $\alpha'$ represents a direct effect. The effect of independent variable on mediator variable is represented by $\beta$ while that between the mediator and the dependent variable is shown by $\gamma$.

Further, Hayes uses a series of Ordinary Least Square regression equations below to sufficient describe a simple mediation model.

$$Y = \kappa_1 + \alpha X + \epsilon_1 \quad (1)$$

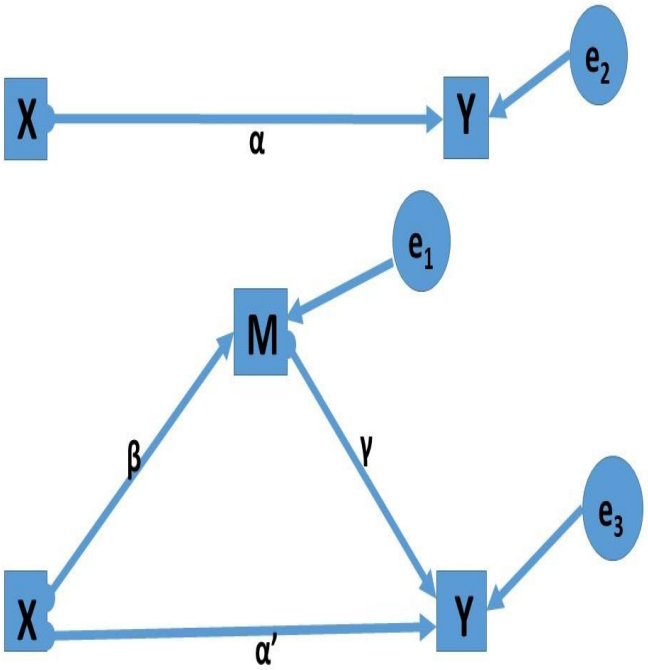


**Figure 1**
*Path diagram: A Simple Mediation Model*

$$M = \kappa_2 + \beta X + \epsilon_2 \quad (2)$$

$$Y = \kappa_3 + \alpha X + \gamma M + \epsilon_3 \quad (3)$$

$\kappa_1$, $\kappa_2$ and $\kappa_3$ shows the intercepts for each of the three equations, while $\epsilon_1$, $\epsilon_2$ and $\epsilon_3$ are respective residuals assumed to follow normal distribution with mean 0 and variance $\delta_1^2$, $\delta_2^2$ and $\delta_3^2$ respectively. This model was adopted in formulating logistic regression model in presence of mediation which was compared to the model used in absence of mediation as per previous studies by (Huberman et al. 2020) to determine the best performing model. The study's findings suggested that the model with mediation was most preferable as compared to the model without mediation

because of its lower AIC values, and higher Mc Fadden $R2$ values as compared to that of the model in absence of mediation.

### 1.1 Main Objective

The main objective of this study is compare the model formulated in presence of mediation to the model in absence of mediation using a Logistic regression model.

### 1.2 Specific Objectives

2 To formulate two Logistic regression models; one in the absence of mediation and another in the presence of mediation.

3 To compare the performance of the two models and evaluate the adequacy of the model fit.

## II. FORMULATION OF MODELS

### 2.1 Formulation of a Logistic Regression Model in the absence of Mediation and Parameter Estimation

For the KENPHIA data, the response variable used was "HIV final result" (HIV positive-1; HIV Negative-2). The mediator variable used was "ever heard of HIV" (Yes-1; No-2). The independent variables; Behavioral variables, Social variables, Demographic variables and Biological variables were assessed using various questions in the survey as follows;

Behavioral variables as "used condom at last sexual encounter in the past 12 months" (Used condom at last sexual intercourse in the past 12 months-1, Did not use condom at last sexual intercourse in the past 12 months -2, No sexual intercourse in the past 12 months-3).

Social variables as "Education level in Kenya" (1 - No primary, 2 - Incomplete Primary, 3 - Complete Primary, 4 - Complete Secondary).

Demographic variables as "Urban Area Indicator" (Urban =1 ; Rural = 2) and Biological variables as "Gender" (Male =1; Female =2).

A sample of $n$ with $n_i$ independent observations was drawn and used from a data set with a total population size of $N$ with $i$ independent observations each defined as $yi$ = 1 if HIV positive or 0 otherwise.

where $i = 1,2,\cdots N$ .

The distribution of $Y_i$ is a Bernoulli and the probability of an individual sampled being HIV positive is $Pr(Y_i = 1) = \pi$ whereas the probability of the sampled individual being HIV negative is $Pr(Y_i = 0) = 1 - \pi$.

In general, the mean of the binary response variable $Y_i$ can be modeled in terms of predictor variable $x_i$ through a linear function given as;

$$E(Y_i) = \beta_0 + \beta_r x_{iR} \quad (4)$$

where $Y_i$ is the response variable, $x_{ir}$ are the explanatory variables and $\beta_r$ are the unknown parameters to be estimated [20]; $i = 1,2,\cdots ,N$ while $r = 0,1,2,\cdots ,R$. The logit transform is equated to the log-odds of the probability of success and to the linear function with multiple predictor variables using the logistic regression model as follows;

$$\begin{aligned} Log(odds) &= ln\left(\frac{\pi}{1-\pi}\right) \\ &= \beta_0 + \beta_1 x_{i1} +, \cdots \beta_R x_{iR} \end{aligned} \quad (5)$$

**Solving Equation 5 by taking anti log and solving for $\pi$**

$$
\begin{aligned}
ln\left(\frac{\pi}{1-\pi}\right) &= \beta_0 + \beta_1 x_{i1} +, \cdots \beta_R x_{ir} \\
&= \sum_{r=0}^{R} \beta_r x_{ir} \; ; \; i = 1, 2, \cdots, n \\
\left(\frac{\pi}{1-\pi}\right) &= \exp^{\sum_{r=0}^{R} \beta_r x_{ir}} \\
\pi &= \exp^{\sum_{r=0}^{R} \beta_r x_{ir}}(1-\pi) \\
\pi(1+\exp^{\sum_{r=0}^{R} \beta_r x_{ir}}) &= \exp^{\sum_{r=0}^{R} \beta_r x_{ir}} \\
\pi &= \frac{\exp^{\sum_{r=0}^{R} \beta_r x_{ir}}}{1+\exp^{\sum_{r=0}^{R} \beta_r x_{ir}}}
\end{aligned} \tag{6}
$$

**The general form of the joint probability distribution (likelihood)for the binary data is given as, [20];**

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \pi^{y_i}(1-\pi)^{1-y_i} \\
L &= \pi^{\sum_{i=1}^{n} y_i}(1-\pi)^{n-\sum_{i=1}^{n} y_i}
\end{aligned} \tag{7}
$$

**Taking natural logs of the likelihood in Equation 7**

$$
\begin{aligned}
\ln L &= \sum_{i=1}^{n} y_i \ln \pi + \left(n - \sum_{i=1}^{n} y_i\right) \ln(1-\pi) \\
&= \sum_{i=1}^{n} y_i \ln \pi - \sum_{i=1}^{n} y_i \ln(1-\pi) + n \ln(1-\pi) \\
&= \sum_{i=1}^{n} y_i \ln\left(\frac{\pi}{1-\pi}\right) + n \ln(1-\pi)
\end{aligned} \tag{8}
$$

**Substituting $ln\left(\frac{\pi}{1-\pi}\right)$ and $\pi$ in Equation 8**

$$
\begin{aligned}
\ln L &= \sum_{i=1}^{n} y_i \left(\sum_{r=0}^{R} \beta_r x_{ir}\right) + n \ln\left(1 - \frac{\exp^{\sum_{r=0}^{R} \beta_r x_{ir}}}{1+\exp^{\sum_{r=0}^{R} \beta_r x_{ir}}}\right) \\
&= \sum_{i=1}^{n} y_i \left(\sum_{r=0}^{R} \beta_r x_{ir}\right) + n \ln\left(\frac{1}{1+\exp^{\sum_{r=0}^{R} \beta_r x_{ir}}}\right) \\
&= \sum_{i=1}^{n} y_i \left(\sum_{r=0}^{R} \beta_r x_{ir}\right) + n \ln(1+\exp^{\sum_{r=0}^{R} \beta_r x_{ir}})^{-1}
\end{aligned} \tag{9}
$$

**Given that $-1\ln(x) = \ln(x)^{-1}$, thus we obtain the log likelihood function which is the logistic regression model in absence of mediation;**

$$\ln L = \sum_{i=1}^{n} y_i \left( \sum_{r=0}^{R} \beta_r x_{ir} \right) - n \ln \left( 1 + \exp^{\sum_{r=0}^{R} \beta_r x_{ir}} \right) \quad (10)$$

Differentiating the log likelihood with respect to parameters $\beta_r$ and solving provides the maximum likelihood estimates of the model.

**Formulation of a Logistic Regression Model in presence of mediation and parameter estimation**

Equations 1, 2 and 3 were used to fit a simple mediation model in Figure 1

This study considered one independent variable, $Y_i$ with multiple covariates, $X_i$. According to (Namazi,2016) mediation analysis mainly looks at decomposing total effect (TE) of the exposure variable into Indirect effect through a mediator and direct effect whose impacts solely comes from the exposure variable.

The mediation effect is indicated by $\alpha$ and $\gamma$ paths while the direct effect by $\alpha'$ path as shown in Figure 1

In this study both Mediation variable, $M$ and dependent variable, $Y$ were Binary variables and the sample size for estimating the parameters for M regression and Y-regression equations were the same. The product of coefficients ($ab$) method was used in this study because of its strength in considering one regression model for the outcome and another regression model for the mediator thus circumventing the model compatibility issue in the difference method (Cheng, 2021). Assuming the conditional mean model of outcome $Y_i$ in Equation 3.

$$g(E(Y \backslash X,M,e_3)) = \kappa 3 + \alpha' X + \gamma M + \epsilon_3 \quad (11)$$

where $g(.)$ is the logit link function, since the outcome is Binary in nature while $\alpha'$ is the exposure effect on the outcome conditional to the effect of the mediator and error term. $\gamma$ represents the relationship between the mediator variable and outcome variable conditional to the effect of the exposure variable and the error term.

In addition, the product method required fitting the mediator model as shown in Equation 2

$$h(E(M \backslash X,e_2)) = \kappa_2 + \beta X + \epsilon 2 \quad (12)$$

Where $h(.)$ is a logit link function given that our mediator variable is Binary and $\beta$ represents the association between the exposure variable and mediator variable conditional on the effects of the covariates and the error term. $\epsilon_2$ and $\epsilon_3$ are independent mean-zero normal errors.

Huberman *et.al (2020)*, states that the total effect which is the expectation of $X$ on $Y$ can further be decomposed through the mediator into direct and indirect effects as follows;

The total effect of $X_i$ on $Y_i$ can be captured in a regression Equation as;

$Y = X\beta + \epsilon$

Where $X = (x_i)$, $\beta = (\beta_0, \beta_1)'$ and $\epsilon$ represents a mean-zero normal error.

The expectation of Y was given as;

$$\begin{aligned} E(Y) &= E(E(Y/M)) \\ &= E(\alpha' X + \gamma M) \\ &= E(\alpha' X + \gamma(\beta X)) \\ &= \alpha' X + \gamma(\beta X) \\ &= X(\alpha' + \gamma\beta) \end{aligned} \quad (13)$$

**The total effect is represented as**

$$\alpha = \alpha' + \gamma\beta \quad (14)$$

**where** $\alpha'$ **and** $\gamma\beta$ represent the direct and indirect effects, respectively Garson, (2013)

The likelihood for a binary data as in this study is given as

$$\begin{aligned} L &= \prod_{i}^{n} \pi^{y_i} (1-\pi)^{1-y_i} \\ &= \pi^{\sum y_i} (1-y_i)^{n-\sum y_i} \end{aligned} \quad (15)$$

**Taking the log of the likelihood function in Equation 16**

$$\begin{aligned}\ln L &= \sum y_i \ln \pi + \left(n - \sum y_i\right) \ln(1 - y_i) \\ &= \sum y_i \ln \pi - \sum y_i \ln(1-\pi) + n \ln(1-\pi) \\ &= \sum y_i \left(\ln \frac{\pi}{1-\pi}\right) + n \ln(1-\pi)\end{aligned} \tag{16}$$

**Similarly substituting for** $\left(\ln \frac{\pi}{1-\pi}\right)$ **and** $\pi$

$$\begin{aligned}\ln L &= \sum y_i \ln \left(\sum_{r=0}^{R} \beta_{mr} x_{ir}\right) + n \ln \left(1 - \frac{\exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}}{1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}}\right) \\ &= \sum y_i \ln \left(\sum_{r=0}^{R} \beta_{mr} x_{ir}\right) + n \ln \left(\frac{1}{1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}}\right) \\ &= \sum y_i \ln \left(\sum_{r=0}^{R} \beta_{mr} x_{ir}\right) + n \ln(1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}})^{-1}\end{aligned} \tag{17}$$

**Given that** $-1\ln(x) = \ln(x)^{-1}$**, thus we obtain our formulated logistic regression model in presence of mediation as;** (18)

**(18)**

$$\ln L = \sum y_i \ln \left(\sum_{r=0}^{R} \beta_{mr} x_{ir}\right) - n \ln \left(1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}\right)$$

Differentiating the model with respect to parameters $\beta_{mr}$ and solving it gives the Maximum likelihood estimates $\beta_{mr}$.

## III. RESULTS & DISCUSSIONS

Kenya Population-based HIV Impact Assessment 2018 survey data was used since it included HIV prevalence of each of the 47 counties and the National HIV prevalence that included for Mandera, Wajir and Garissa counties which were previously excluded from data collected in the KAIS as indicated in the Kenya HIV estimates report (Musyoki, 2017)

### 3.1 Fitting Logistic Regression Model to KENPHIA data set without mediation and parameter estimates

From the KENPHIA data, our response variable was Final HIV Status; 1-HIV Positive, 0-HIV Negative. The predictor variables in the model include "Gender" as the Biological factor whose responses were 1-Male and 2-Female, "Education level in Kenya" as the social factor with responses; 1-No primary, 2-Incomplete primary, 3complete primary and 4-complete secondary), "Urban Area Indicator" as the Demographic variable with responses; 1-Urban, 2-Rural and "Used condom at last sexual intercourse in the past 12 months", as the Behavioral factor with responses 1 - Used condom at last sexual intercourse in the past 12 months 2 - Did not use condom at last sexual intercourse in the past 12 months 3 - No sexual intercourse in the past 12 months.

**The model without mediation was therefore;**

$$\log\left(\frac{\pi}{1-\pi}\right) = 1.863541 + 0.038917B + 0.014316S + 0.013461D - 0.039834b \quad (19)$$

**Table 1**

*Parametric estimates of the fitted regression model to KENPHIA data without Mediation*

| Coefficients | | | | |
|---|---|---|---|---|
| **(Intercept)** | **B** | **S** | **D** | **b** |
| 1.863541 | 0.038917 | 0.014316 | 0.013461 | -0.039834 |
| Residual Deviance: 1215.6 | | | | |
| **AIC: 525.06** | | | | |

The AIC value obtained while fitting the model was **525.06.**

### 3.2 Fitting Logistic Regression Model to KENPHIA data set with mediation and parameter estimates

A mediator variable was introduced in the earlier formulated model

The study assumed that all the individuals tested were exposed to HIV/AIDs. The mediator variable therefore was Ever tested for HIV and responses were; Ever Tested-1, Never tested-2. The logistic regression model with mediation was then fitted as follows;

$$\log\left(\frac{\pi}{1-\pi}\right) = 1.793732 + 0.036647B + 0.018416S + 0.011259D - 0.031591b + 0.047586m \quad (20)$$

**Table 2**

*Parametric estimates of the fitted regression model to KENPHIA data with Mediation*

| Coefficients | | | | | |
|---|---|---|---|---|---|
| (Intercept) | B | S | D | b | m |
| 1.793732 | 0.036647 | 0.018416 | 0.011259 | -0.031591 | 0.047586 |
| Residual Deviance: 1210.1 | | | | | |
| **AIC: 434.01** | | | | | |

Table 4 shows a Positive correlation between all HIV factors in the model and HIV prevalence, except the biological factors which had a negative correlation with HIV prevalence. The value of AIC obtained while fitting the model was 434.01.

### 3.3 Comparison of the performance of the models

The Akaike Information Criteria (AIC) under the model with mediation (434.01), as indicated in Table 2, was lower compared to that in the model without mediation (525.06), as indicated in Table 1. This clearly indicates that the amount of information lost when fitting the model with mediation was less compared to the amount of information lost when fitting the model without mediation; hence, the model with mediation is a better model and fits the data well.

A comparison was also done based on the assessment of model adequacy using the McFadden $R^2$ criterion. A value of 0.4755762 is quite high for McFadden's $R^2$ for the model with mediation, which indicates that the model fits the

data very well and has high predictive power as compared to a value of 0.3593836 for McFadden's $R^2$ for the model without mediation. (Smith & McKenna, 2013; Windmeijer, 1995))

In addition, the comparison of the two models when fitted with KENPHIA data was done using the density curve shown in Figure 2.

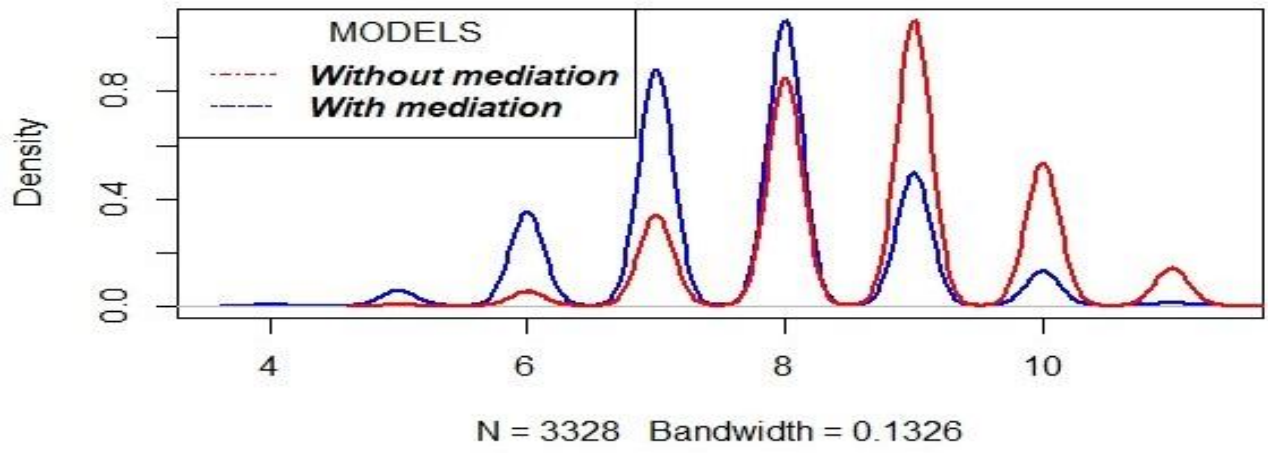


**Figure 2**
*Density Curve of predictor variables for KENPHIA data set in presence and in absence of mediation variable*

The distribution reveals that the model without mediation has several peaks that tend to be negatively skewed with the main mode at 9, while the model with the mediation effect tends towards a normal distribution with the main peak at 8. This indicates that the mediation variable tends to lower the prevalence rate of HIV/AIDS among individuals irrespective of their gender, education level, urban-rural indicator, and condom use, unlike in the model without mediation, where HIV/AIDS prevalence varies with the group in which individuals are in terms of the associated risk factor. Therefore, real data from the KENPHIA survey shows that mediation plays a great role in the prevention of HIV/AIDS in Kenya.

## IV. CONCLUSIONS & RECOMMENDATIONS

### 4.1 Conclusions

The study compared the performance of the model formulated in the absence of mediation and that formulated in the presence of mediation. This was accomplished by comparing the AIC and McFadden's R2 of the two models. The AIC revealed that the model with mediation was of better quality due to the relatively less amount of information lost while fitting the model as compared to the amount of information lost while fitting the model without mediation (Chaba, 2011). The McFadden's $R^2$ by Smith and McKenna (2013) showed that the model formulated in the presence of mediation had a high predictive power since its value was above 0.4; hence, the model fits the data very well as compared to the model without mediation with a McFadden's $R^2$ value below 0.4. A comparison of the two models using the density curves of predictor variables also revealed that the model formulated in the presence of mediation performed better than the model without mediation. Therefore, real data from the KEPHIA 2018 Survey indicates that models formulated in the presence of mediation perform better compared to models without mediation.

### 4.2 Recommendations

Mediation models are very effective in estimating the effect of specific interventions used to control the spread of HIV/AIDS amidst the prevailing HIV-related risk factors. This helps to establish the role played by interventions in HIV prevention instead of generalizing HIV-related risk factors. The study therefore recommends that specific intervention strategies used, such as male circumcision, PrEP, or ART, in HIV/AIDS control be evaluated using mediation models to

establish how effective they are in HIV control. This will help the country channel resources to the specific mediators that are effective and efficient in controlling HIV prevention.

## REFERENCES


Agha, S. (2003). The impact of a mass media campaign on personal risk perception, perceived self-efficacy and on other behavioral predictors. *AIDS Care*, *15*(6), 749-762.

Chaba, L. A. (2011). *Modeling of sti prevalence among hiv-infected adults in HIV care programs in Kenya using logistic regression* (Doctoral Dissertation, University of Nairobi).

Cheng, C., Spiegelman, D., & Li, F. (2021). Estimating the natural indirect effect and the mediation proportion via the product method. *BMC medical research methodology*, *21* (1), 1-20.

Garson, G. D. (2013). *Path analysis*. Asheboro, NC: Statistical Associates Publishing.

Global, A. I. D. S. (2021). *Update. Seizing the moment: tackling entrenched inequalities to end epidemics*. Geneva: UNAIDS; 2020.

Higa, D. H., Crepaz, N., Mullins, M. M., Adegbite-Johnson, A., Gunn, J. K., Denard, C., & Mizuno, Y. (2022). Strategies to improve HIV care outcomes for people with HIV who are out of care. *AIDS*, *36* (6), 853-862.

Huberman, D. B., Reich, B. J., Pacifici, K., & Collazo, J. A. (2020). Estimating the drivers of species distributions with opportunistic data using mediation analysis. *Ecosphere*, *11* (6), e03165.

Irimu, K., & Schwartz, U. (2021). *Reporting HIV/AIDS A guide for Kenyan Journalists* [Internet]. Friedrich Ebert stiftung Coalition of Media Health Professionals.

Joint United Nations Programme on HIV/AIDS. (2013). *Global report: UNAIDS report on the global AIDS epidemic 2013*. Geneva: Joint United Nations Programme on HIV/AIDS.

Karavasilis, G. J., Kotti, V. K., Tsitsis, D. S., Vassiliadis, V. G., & Rigas, A. G. (2005). Statistical methods and software for risk assessment: applications to a neurophysiological data set. *Computational Statistics & Data Analysis*, *49* (1), 243-263.

LaCroix, J. M., Snyder, L. B., Huedo-Medina, T. B., & Johnson, B. T. (2014). Effectiveness of mass media interventions for HIV prevention, 1986–2013: a meta-analysis. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *66*, S329-S340.

Liang, H., & Du, P. (2012). Maximum likelihood estimation in logistic regression models with a diverging number of covariates. *Electronic Journal of Statistics*, *6*(2012), 1838–1846

Mugoya, G. C. T., Aduloju-Ajijola, N., & Dalmida, S. G. (2016). Relationship between Knowledge of Someone Infected with HIV/AIDS and HIV Stigma: A moderated mediation model of HIV knowledge, gender and hiv test uptake. *HIV/AIDS Res Treat Open J.*, *SE*(1),S14-S22. DOI:10.17140/HARTOJ-SE-1-103

Musyoki, R. K. (2017). *HIV and AIDS programmes financing and sustainability in Kenya: a case of the National AIDS Control Council (NACC)* (Doctoral Dissertation, Masters' Thesis, Kenyatta University).

NACC, N. (2018). *Kenya HIV estimates report*. Nairobi, Kenya: NACC.

Namazi, M., & Namazi, N. R. (2016). Conceptual analysis of moderator and mediator variables in business research. *Procedia Economics and Finance*, *36*, 540-554.

Plan, M. C. O. (2017). *Strategic direction summary*. US President's Emergency Plan for AIDS Relief (PEPFAR).

Smith, T. J., & McKenna, C. M. (2013). A comparison of logistic regression pseudo R2 indices. *Multiple Linear Regression Viewpoints*, *39* (2), 17-26.

Srimaneekarn, N., Hayter, A., Liu, W., & Tantipoj, C. (2022). Binary response analysis using logistic regression in dentistry. International Journal of Dentistry, 2022.

UNAIDS, J. (2017). UNAIDS data 2017. Jt. United Nations Program. HIV/AIDS, 1-248.

Windmeijer, F. A. (1995). Goodness-of-fit measures in binary choice models. *Econometric reviews*, *14*(1), 101-116.