Modeling the Effect of Mediation on HIV Prevalence in Kenya using a Logistic Regression Model

Abstract

The control of HIV/AIDS demands different interventions based on various HIV risk factors directly or indirectly affecting HIV prevalence through a mediator variable. There is however limited literature on how these risk factors interact with each other and in turn affect HIV/AIDS prevalence in presence of mediator factors [7]. A logistic regression model formulated in presence of mediation was found to fit both simulted and real data from 2018 Kenya Population-based HIV Impact Assessment (KENPHIA) survey well and had a higher predictive power as compared to the model formulated in absence of mediation. This was accomplished by using Binary logistic regression to fit the models and estimating the model parameters using Maximum Likelihood Estimation in R. Akaike's Information Criterion was used to determine amount of data lost by respective models and McFadden's R^2 to evaluate the adequacy of the model fit.

Keywords: Mediation, Logistic Regression.

1 Introduction

According to UNAIDS Global report [8], 35.3 million people worldwide were living with HIV/AIDS, with 2.2 million being the new infections in 2021. Currently Kenya has a HIV prevalence rate of 4.5% among adults between the ages of 15-49 years and 4.9% among adults between the ages of 15-64 years, [11] &[9]. The variations in HIV prevalence rates cuts across Counties, gender and age [11]. Although the HIV/AIDS epidemic in Kenya has steadily decreased among adults aged 15 to 64 years since 2010, the UNAIDS/WHO AIDS Epidemic Update shows that the actual number of infected individuals is still rising as a result of new infections and longer life expectancy brought

on by the use of anti-retro-viral medications. This calls for further research to be done in order to better understand the causes of HIV sero-positivity and how to control HIV/AIDS in the Nation, [14].

The effective investments done in controlling HIV infections in Kenya [13], has culminated to steady decline in the overall HIV/AIDS prevalence in Kenya, however, there still remain high rates of new infections and differences in the risk of infection, [10]. This could be attributed to the varying effects of interventions used on HIV/AIDS control, either directly or indirectly. Exposure to HIV-related media is one of the most widely used HIV/AIDS control interventions in Kenya, taking into account the social setting in the country. However, due to the numerous HIV risk factors present in the nation, it is nearly impossible to assess the intervention's effect on HIV control and determine whether it is considerably slowing the spread of HIV/AIDS as intended due to its unmeasured indirect effect on HIV/AIDs [6].

In this study, exposure to HIV-related mass media is assumed to be a mediating factor between HIV risk factors and HIV prevalence in Kenya. The biggest challenge however with mass media campaign lies in disseminating accurate, objective, balanced and non-judgmental information on HIV/AIDs to individuals, which implies that the effect of exposure to media in HIV control varies across population groups across the country. A path diagram as indicated in Fig 1 by [4], describes simple relationship between the dependent Y, mediator M and independent X variables. Hayes [4] 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. α is the total effect, $\alpha^{'}$ represents a direct effect and the effect of independent variable on mediator variable is represented by β while that between the mediator and the dependent variable is shown by γ. Further, Hayes uses a series of Ordinary Least Square regression equations below to sufficient describe a simple mediation model.

$$Y = \kappa_1 + \alpha_X + \epsilon_1 \tag{1}$$

$$M = \kappa_2 + \beta_X + \epsilon_2 \tag{2}$$

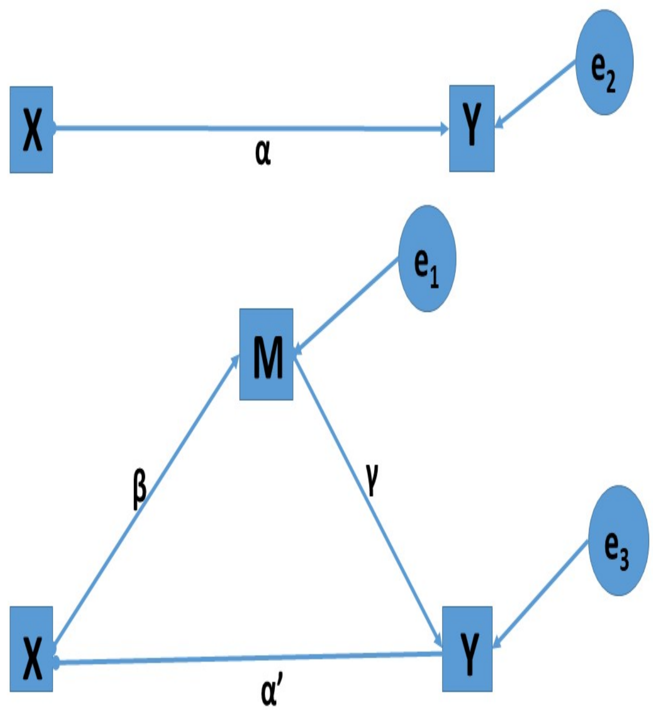

Figure 1: Path diagram: A Simple Mediation Model

$$Y = \kappa_3 + \alpha^{'}_X + \gamma_M + \epsilon_3 \tag{3}$$

κ_1, κ_2 and κ_3 shows the intercepts for each of the three equations, while ϵ_1, ϵ_2 and ϵ_3 are respective residuals assumed to follow normal distribution with mean 0 and variance δ_1^2, δ_2^2 and δ_3^2 respectively. This model was adopted in formulating logistic regression model in presence of mediation which was compared to the model formulated in absence of mediation 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.

2 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 ;

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

where Y_i is the response variable, x_{ir} are the explanatory variables and β_r are the unknown parameters to be estimated [12]. $i = 1, 2, \cdots, N$ 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;

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

Solving Equation 5 by taking anti log and solving for π

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

The general form of the joint probability distribution (likelihood)for the binary data is given as, [12];

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

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** π **in Equation 8**

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

Recall

$-1\ln(x) = \ln(x)^{-1}$ **, thus we obtain;**

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

The log likelihood function in Equation 10 represents the formulated logistic regression model in the absence of mediation. Differentiating the log likelihood with respect to parameters β_r **and solving provides the maximum likelihood estimates in the model.**

Recall that, $\frac{\partial}{\partial\beta_r}\sum_{r=0}^{R}\beta_r x_{ir} = x_{ir}$ **, Since the terms under summation do not depend on** β_r **and assumed to be constants, [3].**

Derivative with respect to β_r **will be**

$$\begin{aligned}\frac{\partial \ln L}{\partial\beta_r} &= \sum_{i=1}^{n} y_i x_{ir} - n\frac{1}{1+\exp^{\sum_{r=0}^{R}\beta_r x_{ir}}}\frac{\partial}{\partial\beta_r}(1+\exp^{\sum_{r=0}^{R}\beta_r x_{ir}})\frac{\partial}{\partial\beta_r}(\sum_{r=0}^{R}\beta_r x_{ir})\\ &= \sum_{i=1}^{n} y_i x_{ir} - n x_{ir}\frac{\exp^{\sum_{r=0}^{R}\beta_r x_{ir}}}{1+\exp^{\sum_{r=0}^{R}\beta_r x_{ir}}} = 0 \qquad (11)\end{aligned}$$

Setting the Equation 11 to zero results to $r+1$ **non-linear equations each having** $r+1$ **unknown parameters that can be solved using iterative process to give maximum likelihood estimates of the models.**

3 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 [4], 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 α and γ paths while the direct effect by $\alpha\prime$ 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 [2]. Assuming the conditional mean model of outcome Y_i in Equation 3.

$$g(E(Y\backslash X, M, e_3)) = \kappa 3 + \alpha\prime X + \gamma M + \epsilon_3 \tag{12}$$

where $g(.)$ is the logit link function, since the outcome is Binary in nature while $\alpha\prime$ is the exposure effect on the outcome conditional to the effect of the mediator and error term. γ 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

$$h(E(M\backslash X, e_2)) = \kappa_2 + \beta X + \epsilon 2 \tag{13}$$

Where $h(.)$ is a logit link function given that our mediator variable is Binary and β represents the association between the exposure variable and mediator variable conditional on the effects of the covariates and the error term. ϵ_2 and ϵ_3 are independent mean-zero normal errors.

Huberman ***et.al*** [5], 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;
$Y = X\beta + \epsilon_1$
Where $X = (x_i)$, $\beta = (\beta_0, \beta_1)'$ and ϵ represents a mean-zero normal error.

The expectation of Y was given as;

$$\begin{aligned} E(Y) &= E(E(Y/M)) \\ &= E(\alpha\prime X + \gamma M) \\ &= E(\alpha\prime X + \gamma(\beta X)) \\ &= \alpha\prime X + \gamma(\beta X) \\ &= X(\alpha\prime + \gamma\beta) \end{aligned} \tag{14}$$

The total effect is represented as

$$\alpha = \alpha\prime + \gamma\beta \tag{15}$$

where $\alpha\prime$ and $\gamma\beta$ represent the direct and indirect effects, respectively [1].

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

$$\begin{aligned} L &= \prod_{i}^{n} \pi^{y_i}(1-\pi)^{1-y_i} \\ &= \pi^{\sum y_i}(1-y_i)^{n-\sum y_i} \end{aligned} \tag{16}$$

Taking the logs of the likelihood

$$\begin{aligned} \ln L &= \sum y_i \ln \pi + \left(n - \sum y_i\right) \ln(1-y_i) \\ &= \sum y_i \ln \pi - \sum y_i \ln(1-\pi) + n\ln(1-\pi) \\ &= \sum y_i \left(\ln \frac{\pi}{1-\pi}\right) + n\ln(1-\pi) \end{aligned} \tag{17}$$

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

$$\begin{aligned} \ln L &= \sum y_i \ln\left(\sum_{r=0}^{R} \beta_{mr}x_{ir}\right) + n\ln\left(1 - \frac{\exp^{\sum_{r=0}^{R}\beta_{mr}x_{ir}}}{1+\exp^{\sum_{r=0}^{R}\beta_{mr}x_{ir}}}\right) \\ &= \sum y_i \ln\left(\sum_{r=0}^{R} \beta_{mr}x_{ir}\right) + n\ln\left(\frac{1}{1+\exp^{\sum_{r=0}^{R}\beta_{mr}x_{ir}}}\right) \\ &= \sum y_i \ln\left(\sum_{r=0}^{R} \beta_{mr}x_{ir}\right) + n\ln(1+\exp^{\sum_{r=0}^{R}}\beta_{mr}x_{ir})^{-1} \end{aligned} \tag{18}$$

Recall that $-1\ln(x) = \ln(x)^{-1}$**, thus we obtain**

$$\ln L \;=\; \sum y_i \ln\left(\sum_{r=0}^{R} \beta_{mr} x_{ir}\right) - n\ln\left(1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}\right) \tag{19}$$

This represents the formulated logistic regression model in the presence of mediation, which is differentiated with respect to parameters β_{mr} **and solved to find the Maximum likelihood estimates** β_{mr}**.**

Recall that,

$$\frac{\partial}{\partial \beta_{mr}} \sum_{r=0}^{R} \beta_{mr} x_{ir} \;=\; x_{ir} \tag{20}$$

Derivative with respect to β_{mr}

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta_{mr}} &= \sum_{i=1}^{n} y_i x_{ir} - n \frac{1}{1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}} \frac{\partial}{\partial \beta_{mr}} (1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}) \frac{\partial}{\partial \beta_{mr}} (\sum_{r=0}^{R} \beta_{mr} x_{ir}) \\ &= \sum_{i=1}^{n} y_i x_{ir} - n x_{ir} \frac{\exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}}{1 + \exp^{\sum_{r=0}^{R} \beta_{mr} x_{ir}}} = 0 \end{aligned} \tag{21}$$

Using the same procedure as mentioned under model in the absence of mediation, $mr{+}1$ **non-linear equations each having** $mr{+}1$ **unknown parameters under this model are solved using iterative process under Newton Raphson Method, and this results into a vector of** β_{mr} **elements.**

4 Data Analysis and Results

Considering a simple linear model involving five covariates including the mediator variable:

$$Y_i = \beta_0 + \beta_1 B_{i1} + \beta_2 S_{i2} + \beta_3 D_{i3} + \beta_4 b_{i4} + \beta_5 m_{i5} + \epsilon_i \tag{22}$$

where $\epsilon_i \sim N(0, \delta_Y^2)$

Fitting Logistic Regression Model to simulated data set without and with mediation and their respective parameter estimates
Let;
P represent HIV status

B represent Behavioral factors
S represent Social Economic factors
D represent Demographic factors
b represent Biological factors
m represent mediation effect

Using R the fitted logistic regression model without mediation is as follows;

$$\log\left(\frac{\pi}{1-\pi}\right) = 0.02980 + 0.04475B + 0.01488S + 0.04608D - 0.04236b \tag{23}$$

whereas the Parametric Estimates as

Table 1: Parametric Estimates

Coefficients				
(Intercept)	**B**	**S**	**D**	**b**
0.02980	**0.04475**	**0.01488**	**0.04608**	**-0.04236**
Residual Deviance: 1385				
AIC: 1395				

The results of the analysis shows that;
There is Positive correlation between Behavioral factors, Social economic factors, Demographic factors and HIV prevalence respectively. However there is Negative correlation between biological factors and HIV prevalence. The value of AIC obtained while fitting the model was 1395.

Fitting Logistic Regression Model to simulated data set with mediation is as follows. The fitted logistic regression model Parametric estimates was as follows;

$$\log\left(\frac{\pi}{1-\pi}\right) = -0.01964 + 0.04248B + 0.01843S + 0.04982D \\ -0.04959b + 0.05256m \tag{24}$$

whereas their parameter estimates as in table 2;

Table 2: Parametric estimates

Coefficients					
(Intercept)	**B**	**S**	**D**	**b**	**m**
-0.01964	**0.04248B**	**0.01843S**	**0.04982D**	**-0.04959b**	**0.05256 m**
Residual Deviance: 1382					
AIC: 1394					

The results shows that there is a Positive correlation between behavioral factors, Social economic factors, Demographic factors and HIV prevalence while biological factors were negatively correlated with HIV prevalence.

The value of AIC obtained while fitting the model was 1394 which is fairly lower, implying that less data was lost when fitting model with mediation as compared to that for the model without mediation.

The lower AIC value and higher McFadden R^2 obtained in the model with mediation indicates that the model formulated in presence of mediation is of good quality and well fits the simulated data than the other without mediation.

The plotted density curve of the model without mediation is steeper compared to the model with mediation. This implies that mediator variable plays a big role in lowering the HIV prevalence as shown in Fig 2.

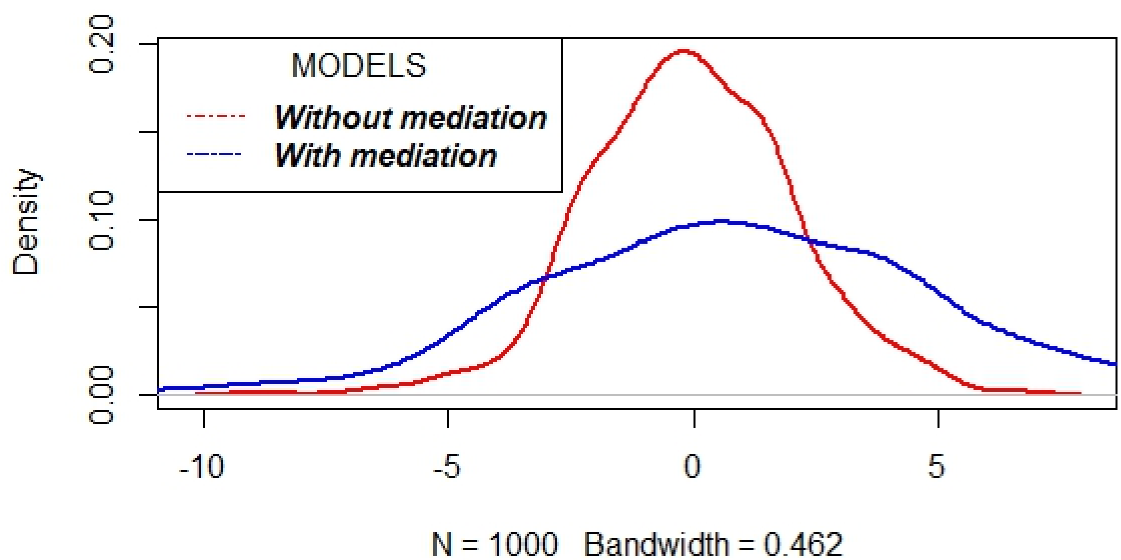

Figure 2: Density Curve of predictor variables using simulated data in presence and in absence of mediation

5 Kenya Population-based HIV Impact Assessment (KENPHIA) 2018 survey data

This survey was carried out with an aim of building on the previously conducted Kenya AIDS Indicator Survey (KAIS) surveys. The new features in the survey 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 [9].

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, 3-complete 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 logistic regression model without mediation was then fitted as follows;

$$\log\left(\frac{\pi}{1-\pi}\right) = 1.863541 + 0.038917B + 0.014316S + 0.013461D - 0.039834b \tag{25}$$

Table 3: 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				

Table 3 shows a Positive correlation between Behavioral factors, B **and HIV prevalence, a positive correlation between Social economic factors,** S **and HIV prevalence, and a Positive correlation between Demographic factors,** D **and HIV prevalence. The table further shows a Negative correlation between biological factors,** b **and HIV prevalence. The AIC value obtained while fitting the model was 525.06.**

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

To fit the model with mediation, a mediator variable was introduced in the earlier formulated model under equation 25

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;

$$\log\left(\frac{\pi}{1-\pi}\right) = 1.793732 + 0.036647B + 0.018416S$$
$$+0.011259D - 0.031591b + 0.047586m \tag{26}$$

Table 4: 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 Behavioral factors and HIV prevalence, a positive correlation between Social economic factors and HIV prevalence and a Positive correlation between Demographic factors and HIV prevalence. There is also a Negative correlation between biological factors and HIV prevalence and the value of AIC obtained while fitting the model was 434.01.

Therefore,introduction of mediation in the model lowered the effect of each parameter on HIV prevalence hence overally bringing the HIV prevalence levels down.

Similarly, the model with mediation had better quality with less data lost and had a higher predictive power due to higher McFadden's R^2 obtained as compared to the model without mediation. The plotted density curve in figure 3 indicates that 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, using both simulation and real data from KENPHIA survey shows that mediation plays a great role in prevention of HIV/AIDS in Kenya.

6 Summary of results and conclusion

The study aimed at modeling the effect of mediation on HIV/AIDs Prevalence in Kenya using logistic regression model. In both cases the model with mediator variable present was found to be of better quality with a high predictive power as indicated by the values of AIC and McFadden R^2 values presented in the table 5.

7 Recommendation for Further Research

As a result of complexities in the relationship between HIV related risk factors and the prevailing socio economic, demographic and cultural background of individuals in Kenya, there is need to estimate the impact of other inter-

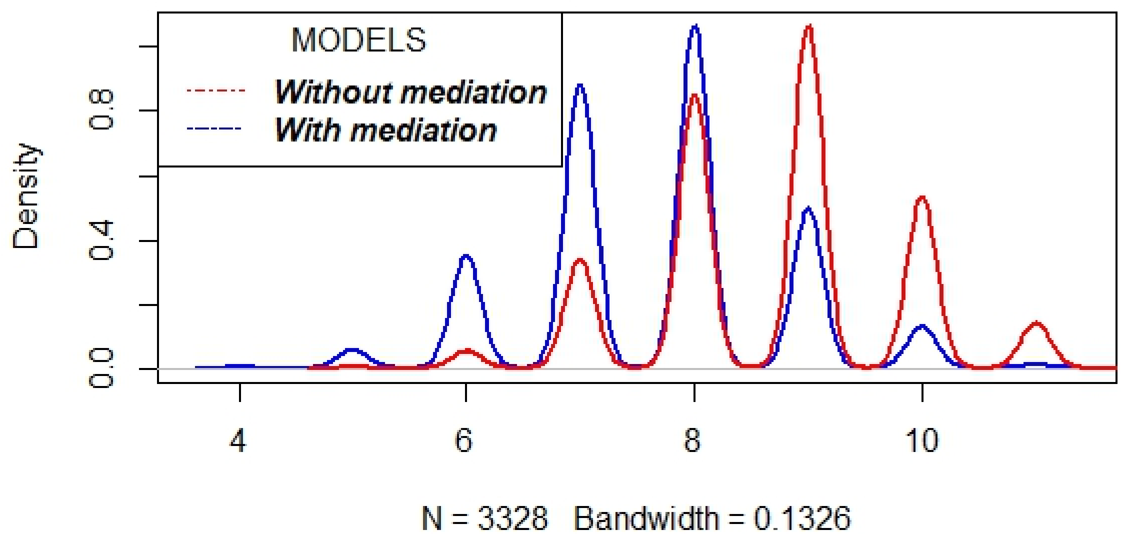

Figure 3: Density Curve of predictor variables for KENPHIA data set in presence and in absence of mediation variable

Table 5: Selection process of the best model using simulated and KENPHIA data

Model parameter	Estimated parameter effect	AIC Value	McFadden R^2 Value
Simulation Without mediation			
B	0.04475		
S	0.01488	1395	0.001048959
D	0.04608		
b	- 0.04236		
Simulation With mediation			
B	0.04248		
S	0.018413	1394	0.003113601
D	0.04982		
b	- 0.04959		
m	0.05256		
KENPHIA Without mediation			
B	0.038917		
S	0.014316	525.06	0.3593836
D	0.013461		
b	- 0.039834		
KENPHIA Without mediation			
B	0.036647		
S	0.018416	434.01	0.4755762
D	0.011259		
b	- 0.031591		
m	0.047586		

ventions used in HIV prevention in Kenya. This will help the country rightly channel resources to interventions that have great impact on HIV prevention, hence this study recommends further work to be carried out on the prevalence of HIV in Kenya in presence of other mediation factors in the fight against HIV AIDS.

References

[1] **Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis.** *American sociological review,* **37-47.**

[2] **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.**

[3] **Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. Available at czep. net/stat/mlelr. pdf, 83.**

[4] **Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium.** *Communication monographs,* **76 (4), 408-420.**

[5] **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.**

[6] **Irimu, K., & Schwartz, U. (2021). Reporting HIV/AIDS A guide for Kenyan Journalists [Internet]. Friedrich Ebert stiftung Coalition of Media Health Professionals; 2003.**

[7] **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.**

[8] **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.*

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

[10] **National AIDS Control Council, Kenya (NACC). (2005). Kenya National AIDS Strategic Plan 2005/6–2009/10 (KNASP).**

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

[12] **Stephenson, B., Cook, D., Dixon, P., Duckworth, W., Kaiser, M., Koehler, K., & Meeker, W. (2008). Binary response and logistic regression analysis. available at:¡ a href=" http://www. stat. wisc. edu/mchung/teaching/MIA/reading/GLM. logistic. Rpackage. pdf"¿ http://www. stat. wisc. edu/mchung/teaching/MIA/reading/GLM. logistic. Rpackage. pdf¡/a¿(last access: 30 August 2014).**

[13] **World Health Organization. (2007). Joint United Nations Program on HIV/AIDS.** *HIV/AIDS programme. Strengthening health services to fight HIV/AIDS. Guidance on provider-initiated HIV testing and counseling in health facilities. Geneva, Switzerland: WHO.*

[14] **World Health Organization. (2010).** *AIDS epidemic update: December 2009.* **WHO Regional Office Europe.**