JBE Journal of Biostatistics and Epidemiology
J Biostat Epidemiol. 2025;11(1): 36-49

ORIGINAL ARTICLE
DOI: https://doi.org/10.18502/jbe.v11i1.19316
Received October 19, 2024; Accepted December 17, 2024;
Published online March 15, 2025.

Estimation of HIV Prevalence among Women in Kenya in the Presence of Mediation Using Latent Trait Analysis

Tilence **Mola**, Drinold **Aluda Mbete***, Job **Sirengo**

Department of Mathematics, School of Natural Science, Masinde Muliro University of Science and Technology, Kakamega, Kenya.

ABSTRACT

Introduction: Estimating prevalence in cause-effect relationships where the mediator variables are assumed to be latent is not usually easy. However, the use of proper indicators and statistical model can make the measurement and use of such constructs easy.

Methods: Structural Equation Modeling makes it possible to analyze simultaneously both the relationship between the latent variable and the links between the latent variable and their indicators. The 2018 Kenya AIDS Indicator Survey data was used to validate the model developed. The maximum likelihood was used to estimate the model parameters. The findings of the study were, there is a relationship between education attainment and knowledge /awareness of HIV/AIDS.

Results: The results further shows that education levels are not associated with HIV prevalence after controlling for a number of socio-demographic characteristics and behavioral factors.

Conclusion: These findings can inform policy makers in formulation of appropriate HIV/AIDS management (policies) and intervention strategies aimed at reducing HIV/AIDS prevalence that has remained a challenge in many developing countries.

Key words: HIV prevalence; Continuous latent mediator; Latent trait analysis; Observed indicators

*Corresponding Author:
dmbete@mmust.ac.ke

INTRODUCTION

In 2018, an estimated 35.0 million people were living with HIV worldwide where Sub-Saharan Africa (SSA) accounts for 71% of the population yet is home to only 12% of the global population. According to UNAIDS report,[1] African countries account for almost (80%) of all people living with HIV, for instance, South Africa (25%), Nigeria (13%), Mozambique (6%), Uganda (6%), Tanzania (6%), Zimbabwe (6%), Kenya (6%), Zambia (4%), Malawi (4%) and Ethiopia (3%). The report further indicates that, trends in new HIV infections across countries in SSA have shown a decline by more than 33% from an estimated 2.2 million in 2005 to 1.5 million in 2013, but still remain high.[2]

Transmission of HIV in SSA is majorly through sexual contact and has remained a challenge to the possibility of an AIDS free generation, despite there being many initiatives to prevent it. Globally, 15% of women living with HIV are aged 15-24 years, of whom 80% live in SSA.[1, 3] The Kenya HIV estimate report of 2020 shows that the nation adult population aged between 15-49 years have prevalence estimated at 4.9% with higher prevalence among women (5.2%) than men (4.5%). However, the spectrum results show a continued decline over a period of time since 2010. The trend indicates that the prevalence peaked at 10-11% in the mid-1990s, declined to about 6% in 2006 and has been relatively stable for several years.

Despite the progress by Kenya in advancing towards the national target of achieving an AIDS free generation as reflected in the Kenya HIV Estimates Report 2020,[2] much remains to be done to prevent new HIV infection. The report recommended that in order to attain an AIDS-free generation, it is crucial that the focus is geared towards the young women. This is because, in over 70% of all new HIV infections that occur, young women bear a high rate of infection 39%.[1] They also acquire HIV infection 5-7 years earlier than their male peers. According to the report, they are driven by behaviors such as; early sexual debut, gender based violence, inconsistent condom use with multiple partners, low acceptability of condom use in cohabiting couples among others.

The Kenya HIV Estimate report[2] suggests that, prevention efforts must focus on broad social movements that contribute to safer sex behaviors in young women and extend to the general population with increased vulnerability to HIV. The report further explains that, the goal of HIV intervention research is to develop interventions that encourage participants to reduce or eliminate sexual or social behaviors that put themselves or others at increased risk for HIV infection.

The study sought to establish how the education levels of Kenyan women aged between 1564 years affect their HIV status through unobserved sexual behaviour patterns measured by three categorical indicators (use of condoms, non-marital sexual relationship and abstinence). The survey data used contained data measured on different metric scales, that is; binary outcome (HIV status), continuous latent mediator (sexual behaviour patterns), and categorical predictor variable (education levels). Such a dynamic and complex multivariate model calls for a powerful statistical analysis tool/ technique. Structural Equation Modeling (SEM), and more specifically, the latent trait analysis (LTA) was used.

Latent Trait Analysis

A latent variable is a type of a variable that cannot be observed or measured directly. To measure it, researchers capture indicators that are believed to accurately represent it. It is assumed that, the responses on the indicators are the result of an individual's position on the latent variable(s).[4] In this study, the latent variable i.e. sexual behavior patterns have three indicators (i.e. use of condoms, abstinence and one-sexual partner relationship) that are believed to accurately represent and measure it.

Latent variables are grouped according to whether they are manifest, categorical or continuous. If the latent variables are categorical, latent class analysis (LCA) is used, while latent trait analysis (LTA) is used when the latent variable is continuous. Most mediation studies involve latent traits since one can almost never measure the variables of primary interest directly, hence it is important to consider statistical models that include latent traits. A latent variable model is a statistical model that relates a set of observable (manifest) variables to a set of latent variables. It consists of a general class of models suitable for the analysis of multivariate data as explained by.[5-6] Therefore, Latent trait analysis (LTA) is a multivariate regression model that links categorical or continuous responses to unobserved continuous variables.

This study assumes categorical predictor variables, continuous latent mediator variables with categorical indicators and a binary outcome. As such, latent trait analysis (LTA) approach is used (a) to aggregate large number of variables into a model (reduce the dimensionality of data) that represent an underlying concept. Thus, conclusions in complex situations are easily drawn,[6] (b) since they are powerful and flexible method for identifying and understanding unobserved groups in a population, and (c) link observable data in the real world to symbolic data in the modeled world.[7]

This study therefore aims to establish the strength of the association between the categorical covariate(s) and the binary response through latent continuous mediator variable(s) using the structural equation modeling. The direct, indirect and total effects were examined in a single model. Specifically, the sexual behavior patterns (intermediate variable) were used to explain how or why an independent variable (education level) affects the outcome HIV (status). Furthermore, the model parameters were estimated using the maximum likelihood (ML) estimation approach.

Literature Review

Many scholars are interested in understanding the process by which an independent variable affects a dependent variable either directly or indirectly through a mediator. Researchers have tested for mediation when all the variables are continuous, but a definitive answer had been lacking as to how to analyze the data when predictor variables are categorical, latent mediator variables are continuous and with categorical indicators and a binary outcome. Survey data in economic, social and medical fields contains variables which are measured on different scales. They may be on binary scale, categorical scale (nominal or ordinal), metric scale (discrete or continuous) or a combination of the above. Researchers have proposed various solutions to working with categorical variables or a mix of categorical and continuous variables in mediation. For example,[8] focused on predictor X and allowed it to be categorical, not just binary and suggested that it would be ideal to allow for categorical M and Y as well. Other researchers make the assumption that, while a manifest variable may be discrete, the underlying construct is continuous. For example,[9] and [10] proposed modifications to structural equation modeling, of which the mediation form is a special case, for categorical variables. Muthen[9] assumed categorical predictor variables, continuous latent variable with observed categorical data arising through a threshold step function. His estimation procedure was based on generalized least squares, assumed normal distributions, and recommended a large sample size. Motivated by Muthen's study, the current study will use latent trait analysis to test for mediation and report findings when predictor variables are categorical, latent mediator variables are continuous with categorical indicators and a binary outcome.

The study by Krueger[11] used the latent trait model to understand the nature of alcohol problems. They first developed a two-parameter latent trait model and fitted it to the data on 110 alcohol problems reported during in-person interviews of a large sample of middle-aged men from the general population. The validity of the model was assessed by Exploratory Factor Analysis (EFA) and used MicroFACT computer program to analyse the data. Their findings demonstrated how the two parameters map alcohol problems along a continuous scale of variation and how latent trait methods can provide an empirical way to understand the alcohol problems. The study only assessed the direct effect and never provided the indirect effect results which led to inadequate support of the hypothesized model by the findings.

Li[12] in his study examined the potential for extending the Trans-theoretical Model (TTM) to the area of complementary and alternative medicine (CAM) use. Based upon the literature, a hypothesized structural equation model of readiness of general CAM use was established. The research was conducted among 518 Australian university students. The hypothesized structural model showed a satisfactory degree of fit to the observed data, the finding provided quantitative evidence of the applicability of the TTM to study the readiness of CAM use. However, the hypothesized theoretical model was minimally supported by the findings just like the study by Krueger above. This is due to lack of the mediating effect which is an indirect effect. These two studies by Krueger and Jie show the application of SEM without mediation. The indirect effect is the part of the effect of the independent variable that is mediated, or transmitted, by another variable. By assessing indirect effects as well as direct effects, the total effects of each construct on the dependent variable is more thoroughly evidenced. For this reason, our study developed a model with a mediator variable to allow assessment of both direct effects and indirect effects. While the basic logic analyses have stood the test of time, there has been quite a bit of research aimed at improving the methodology.

In particular, MacKinnon[13-15] and his colleagues have made tremendous contributions to improving both the accuracy and precision in mediation procedures. Mediation modeling has been extended to incorporate measurement errors which when overlooked affect the reliability and validity of observed scores. Compared to simple path or regression analysis, structural equation models are superior and latent variable approach to mediation is more powerful than the other approaches. Conway et al[16] compared the latent trait model, the latent class model, and factor mixture models. Their aim was to look for the final model that would provide a good fit to the data. According to the hypothesis, they concluded that, the latent trait model clearly provided the superior fit to the data when compared to the various latent class and factor mixture models for this reason our study employed the LTA method to report the findings.

Model Formulation

The latent trait model is divided into two
i) Structural Model
ii) Measurement Model
The structural model equation is represented by the matrix below

$$\eta=\beta\eta+\Gamma x+\zeta \qquad (1)$$

From the structural model equation above, η is a vector representing dependent variables (*M* and *Y*), *X* is a vector representing independent variables (*X*) and ζ is a vector representing random errors ϵ_1 and ϵ_2 respectively. The parameters β and Γ are as defined above.

The measurement model gives the relationship between observed dichotomous variables and unobserved latent variable. It describes how each latent variable is defined via the manifest variables, and provides information about the validity and reliability of the structural model. The measurement model is given by;

$$y=\Lambda_y x+\varepsilon \qquad (2)$$

Where; $\Lambda=(\beta,\eta,\Gamma,\gamma_1,\gamma_2,\sigma_\eta,\sigma_y)$. Note that, y is a vector of the observed endogenous variables and ε is a vector of errors. Λ_y is the matrix of the structural coefficients between the observed variables and the latent variables.

Estimation of Indirect Effect

The indirect effects of a variable are mediated by at least one intervening variable. The sum of the direct and indirect effects is the total effects. In our model the indirect effect is illustrated by the influence of *X* on *Y* through the intervening variable η. It is expected that, one unit change in *X* leads to an expected β change in η. Thus β change in η should lead to an expected γ_1 change in *Y*. Therefore, the indirect of *X* on *Y* is $\beta\gamma_1$. Figure 1 b represents an indirect as well as direct effects.

The model assumes not just a direct effect from the predictor *X* to the outcome *Y* but also an effect of *X* on the mediator *M*.[17] As such both *Y* and *M* can be written as linear functions of their predictors and normally distributed error terms. The equations for *Y* and *M* that are used to define the indirect effect can be combined to obtain the total effect of *X* on *Y* under model 2 (Figure 2). From the structural model note that ϵ_1 represents the error term for *M* and ϵ_2 represents the error term for *Y*. Assume that ϵ_2 is independent of *M* and of ϵ_1 and let var $\epsilon_2=\sigma_Y^2$ and var $\epsilon_1=\sigma_1^2$.
Using the specification of *Y* and *M* under Model 2 and under the assumption that *Y* and *M* have a

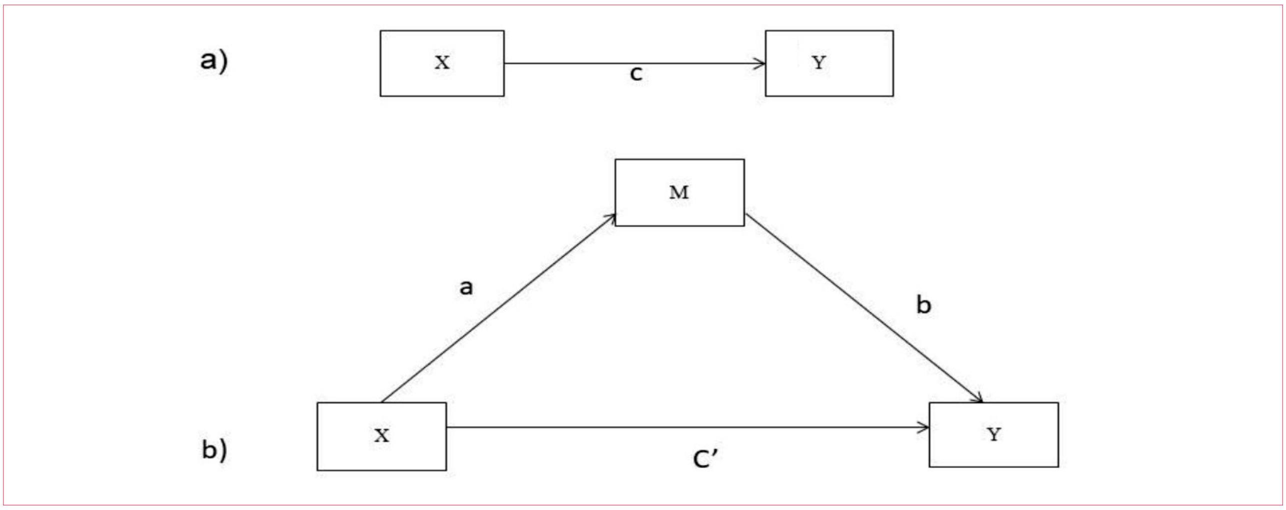

Figure 1. Path diagram

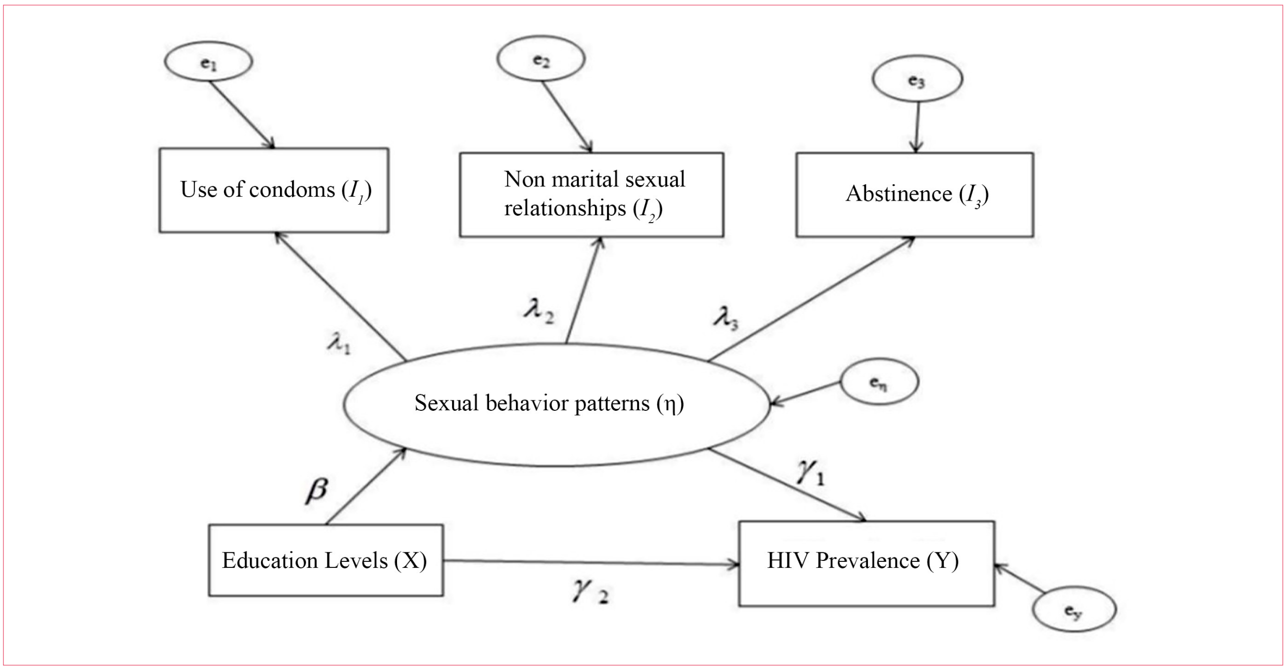

Figure 2. Latent trait model

multivariate normal distribution: Based on the multivariate normal distribution, the probability density function of Y and M can be written as,

$$f\left(Y, \frac{M}{X}, \gamma_1, \gamma_2, \beta\right) = (2\pi)^{-n}(\sigma_Y^2\sigma_M^2)^{\frac{-n}{2}}\exp$$
$$\frac{-\nu_{11}}{2}\sum_{i=1}^{n} a_i^2 - \nu_{12}\sum_{i=1}^{n} a_i b_i - \frac{\nu_{22}}{2}\sum_{i=1}^{n} b_i^2 \tag{3}$$

Where, $a_i = y_i - (\gamma_1 + \gamma_2 \beta)\, x_i$ and $b_i = M_i - \beta X_i$

The probability density function of Y and M can be viewed equivalently as the joint likelihood of γ_1, γ_2, and β.Therefore, from the likelihood the maximum likelihood estimates (MLEs) for γ_1, γ_2, and β can be obtained. Once the MLEs have been calculated, they can be substituted into the likelihood to obtain estimated or profile likelihoods for the parameters of interest, namely γ_2 and β. The MLEs is found by maximizing the log of the likelihood,

$$L\left(\gamma_1,\gamma_2,\frac{\beta}{X},M,Y,,\sigma_Y^2,\sigma_M^2\right) = \\ -\,\mathrm{nln}\,2\pi - \frac{n}{2}\ln\sigma_Y^2\,,\sigma_M^2 - \frac{v_{11}}{2} \\ \sum_{i=1}^{n} a_i^2 - v_{12}\sum_{i=1}^{n} a_i b_i - \frac{v_{22}}{2}\sum_{i=1}^{2} b_i^2 \tag{4}$$

Next, the derivative of the log likelihood with respect to the variable that is being maximized must be set equal to zero. Then, the value for the variable that solves the equation is considered the maximum likelihood estimate. The MLE for γ_1 is obtained as shown in

$$\frac{\partial l}{\partial\gamma_1} = v_{11}\sum_{i=1}^{n} a_i x_i + v_{12}\sum_{i=1}^{n} b_i x_i$$

$$= \frac{1}{2\sigma^2 Y}\sum_{i=1}^{n}\,[\mathrm{y} - (\gamma_1+\gamma_2\beta)x_i]x_i - \frac{\gamma_2}{\sigma^2 Y}\sum_{i=1}^{n}\,(M_i - \beta x_i)x_i$$

$$= \frac{1}{2\sigma^2 Y}\sum_{i=1}^{n} y_i x_i - \frac{\gamma_1}{2\sigma^2 Y}\sum_{i=1}^{n} x_i^2 - \frac{\gamma_2\beta}{2\sigma^2 Y}\sum_{i=1}^{n} x_i^2 - \frac{\gamma_2\beta}{2\sigma^2 Y}\sum_{i=1}^{n} m_i x_i + \frac{\gamma_2\beta}{2\sigma^2 Y}\sum_{i=1}^{n} x_i^2$$

$$= \frac{1}{\sigma^2 Y}\sum_{i=1}^{n} y_i x_i - \frac{\gamma_1}{\sigma^2 Y}\sum_{i=1}^{n} x_i^2 - \frac{\gamma_2}{2\sigma^2 Y}\sum_{i=1}^{n} m_i x_i$$

$$= \frac{1}{\sigma^2 Y}\sum_{i=1}^{n}\,(y_i - \gamma_2 m_i)x_i -= \frac{\gamma_1}{\sigma^2 Y}\sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n}\,(y_i - \gamma_2 m_i)x_i}{\sum_{i=1}^{n}\,x_i^2} \tag{5}$$

Using a similar process, we can obtain the MLE for γ_2

$$\frac{\partial l}{\partial \gamma_2} = v_{11}\sum_{i=1}^{n} \beta a_i x_i + v_{11}\sum_{i=1}^{n} a_i b_i - v_{12}\sum_{i=1}^{n} \beta b_i x_i - v_{12}\sum_{i=1}^{n} b_i^2$$

$$= \frac{\beta}{\sigma^2 Y}\sum_{i=1}^{n} (a_i - \gamma_2 b_i)x_i + \frac{1}{\sigma^2 Y}\sum_{i=1}^{n} (a_i - \gamma_2 b_i)b_i$$

$$\frac{1}{\sigma^2 Y}\sum_{i=1}^{n} (a_i - \gamma_2 b_i)(\beta x_i - b_i)$$

(6)

Replacing a_i and b_i in the equation above, the following iis obtained,

$$\frac{1}{\sigma_Y^2}\sum_{i=1}^{n} y_i - \gamma_1 x_i - \gamma_2 m_i = 0$$

(7)

Recall that

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{n} (y_i - \gamma_2 m_i)x_i}{\sum_{i=1}^{n} x_i^2}$$

(8)

Then substitute in the MLE for γ_1 into the equation to obtain γ_2, the MLE for γ_2, which will not depend on γ_1

$$\sum_{i=1}^{n} y_i m_i - \sum_{i=1}^{n} (y_i - \gamma_2 m_i)x_i \sum_{i=1}^{n} x_i \sum_{i=1}^{n} - \hat{\gamma_2}\sum_{i=1}^{n} m_i^2 = 0$$

$$\sum_{i=1}^{n} y_i m_i - \frac{\sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} x_i m_i}{\sum_{i=1}^{n} x_i^2} + \hat{\gamma_2}\frac{\sum_{i=1}^{n} x_i m_i}{\sum_{i=1}^{n} x_i^2} - \hat{\gamma_2}\sum_{i=1}^{n} m_i^2 = 0$$

$$\sum_{i=1}^{n} y_i m_i - \frac{\sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} y_i m_i}{\sum_{i=1}^{n} x_i^2}\hat{\gamma_2}\left(\frac{(\sum_{i=1}^{n} x_i m_i)^2}{\sum_{i=1}^{n} x_i^2} - \sum_{i=1}^{n} m_i^2\right) = 0$$

$$\frac{\sum_{i=1}^{n} y_i m_i - \frac{\sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} x_i m_i}{\sum_{i=1}^{n} x_i^2}}{\sum_{i=1}^{n} m_i^2 - \frac{(\sum_{i=1}^{n} x_i m_i)^2}{\sum_{i=1}^{n} x_i^2}} = \hat{\gamma_2}$$

(9)

which simplifies to

$$\hat{\gamma}_2 = \frac{\sum_{i=1}^{n} y_i m_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} x_i m_i}{\sum_{i=1}^{n} m_i^2 \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i m_i)^2} \quad (10)$$

RESULTS AND DISCUSSION

The general Linear regression analysis

The correlation test results between independent variables and dependent variable are shown in Table 1.

Table 1. Correlation between dependent and independent variables

Model Estimate results	Estimate	S.E	Est/S.E	r-values	P-value
HIV Outcome	0.625	0.054	11.656	0.8461	<0.001
Education level	0.581	0.059	9.855	0.8893	<0.001

The results show that the independent variable is correlated with the dependent variable. This is indicated by p-values that are less than α, where α value is 0.05. Therefore, there exist a relationship between independent (education level) variables with dependent variable (HIV outcome). Further, the estimates of the standard factor loadings are as shown by the model results in Table 2.

Table 2. Estimates of the standard factor loadings

LAMBDA	Sexual Behaviour patterns	HIV outcome	Education levels
HIV outcome	0.0021	1.000	0.000
Use of Condoms	0.001	0.000	0.000
Abstinence	0.192	0.000	0.000
Non Marital Sexual relationships	0.002	0.000	0.000
Education levels	0.000	0.000	1.000

The results indicate that the HIV outcome is strongly correlated with the mediator (Sexual Behaviour Pattern). This is shown by low values of the indicators representing the mediator, that is, 0.001 for use of condoms, 0.192 for abstinence and 0.002 for Non-Marital Sexual relationships. The analysis to test the goodness of fit of the measurement model was also carried out so as to assess the extent to which the latent variable was represented by the indicators (abstinence, use of condoms and non-marital sexual behaviour). A Chi-Square goodness of fit test was used to determine whether or not a categorical variable follows a hypothesized distribution. From Table 1, the results show p-values <0.05 which means that the data provide sufficient evidence against the null hypothesis, so it was concluded that the observed data follows a different distribution than the theoretical one. The overall fit of the model parameters had a Chi-square estimate of 21.695 with a p-value of 0.00002.

To compare two or more models i.e. measurement model and theoretical model, the Akaike Information Criterion (AIC) was used where AIC with smaller value of standardized estimates represents a better fit of the hypothesized model. The results indicated that the measurement model had AIC = 1.178 and theoretical model had AIC = 2.695. The measurement model has a better fit compared to theoretical model since it has the lowest Akaike Information Criterion.

Measurement Model

The measurement model included one latent factor (Sexual Behaviour Patterns) and three observed variables (Use of Condoms, Abstinence and Non-Marital Sexual behaviours). This model was estimated using the data 2018 KAIS which was a complete data since listwise deletion for missing data was used. Univariate skewness values ranged from -0.07 to -1.12 and univariate kurtosis values ranged from -0.79 to 0.91 confirming that the variables were indeed normally distributed[18] and that no special estimators to address non normality were necessary. The overall fit indices and parameter estimates, taken directly from the output, are shown in Table 3 and 4. The tables presents results in the form of regression equations, where the first set of equations is the unstandardized parameter estimates, and the second set is the standardized parameter estimates.

Table 3. Overall fit and significance of model parameters

RMSEA	Estimate
Chi-square	21.695 (2df)
Probability value for the Chi-square statistic	0.00002
Chi-square for this ML solution	20.686.
Comparative fit index (CFI)	0.907
standardized RMSR	0.044
Estimate	0.068
90 percent CI	0.345 0.419
Probability RMSEA	0.05 0.000

Table 4. Standardized parameter estimates

Sexual Behavior Patterns BY	Estimate	S.E	Est/S.E	P-value
Use of Condoms	0.982	0.000	999.0	<0.0001
Abstinence	0.933	0.039	23.74	<0.0001
Non-Marital Sexual Behaviors	0.880	0.038	23.057	<0.0001

From the results in Table 3, the overall significance of model parameters shows that the estimates of the model parameters were well represented by its indicators. An initial test of the measurement model revealed a very satisfactory fit to the data: χ^2 (21.685, N = 5000) = 113.22, p<.001; RMSEA = 0.055; SRMR = 0.044; and CFI = 0.907. All the factor loadings for the indicators on the latent variables were significant (p< 0.001), indicating that the latent factor was well represented by its indicators.

From the results in Table 4 abstinence had a higher estimate 0.933 as compared to Non-Marital Sexual behaviours with an estimate of 0.880 when use of condoms was controlled.

Structural Model

In Table 5 and 6, output of model parameters is shown. From Table 5, Mplus provides four columns of output for each estimated parameter. The first column depicts the unstandardized coefficient; the second column, the standard error; the third column, the z-score demonstrating the significance of each parameter and the fourth column the standardized solution.

Table 5. Parameter estimates of structural parameters

Intercepts	Estimate	S.E	Est/S.E	STD
HIV Outcome	0.301	0.067	4.473	0.142
USE of Condoms	0.385	0.074	5.219	0.221
Abstinence	0.075	0.087	0.861	0.389
Non Marital Sexual Behaviors	0.126	0.084	1.495	0.135

Table 6. Fit Indices for structural model

RMSEA	Estimate
Estimate	0.071
90 percent CI	0.345 0.419
CFI	0.96
TLI	0.710
Probability RMSEA	0.05 0.000

From Table 5, Abstinence has a higher significance 0.861 as compared to other indicators, while non marital sexual behavior has a lower standardized coefficient. The results further show that the disturbance terms (errors in prediction) were correlated among each construct, it shows there was a significant indirect effect (mediated) of the sexual behavior pattern on HIV outcome.

In Table 6 the model shows appropriate fit according to all indices. They provide the estimates of the latent trait factor with a mean of (94.342) and variance of (101.86) as well as the estimate of the measurement error variances under residual variances values ranging between 9.389 to 11.958 (Figure 3).

Direct and Indirect effect

The comparison between the Direct model and the Indirect models is shown in Table 7. From the results in Table 7, the Indirect model has a smaller standardized value (0.0783) as compared to the Direct model (0.1442). The total effect is the combination of both the direct and indirect effect. It also has a lower estimate of 0.066 as compared to the direct one with an estimate of 0.077. Looking at their

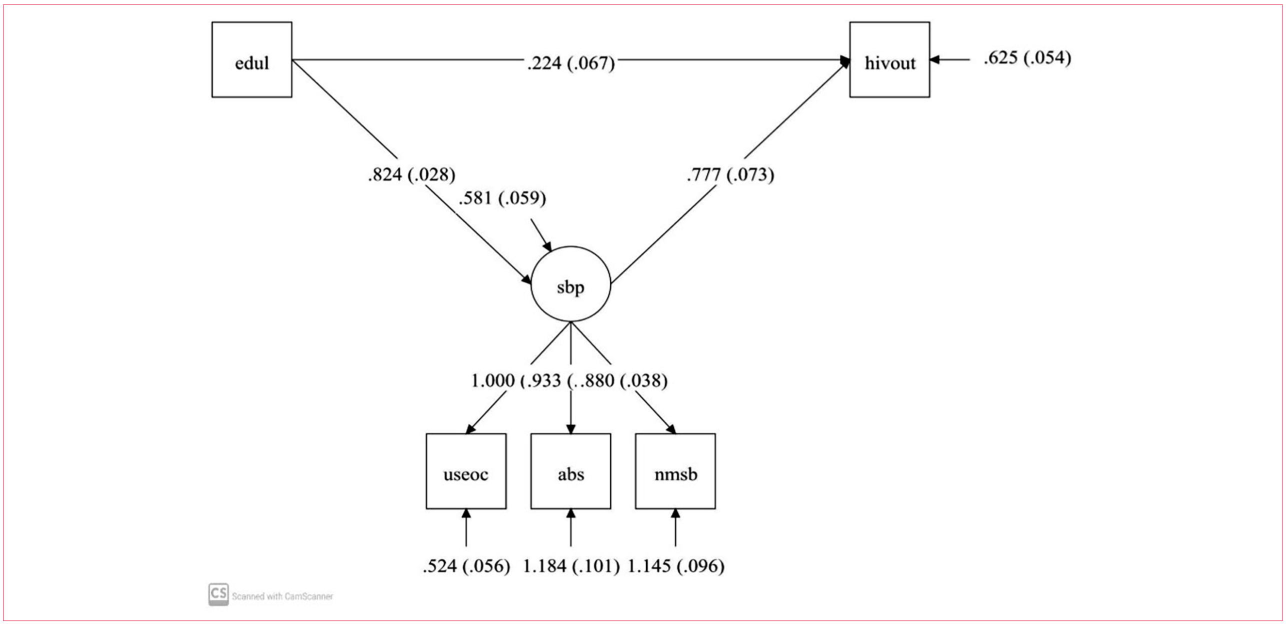

Figure 3.Output diagram

estimates the indirect effect is statistically significant but the direct is not. The reliability of the mean from the results can be deduced from the Standard Error results which shows that small values of SE indicates that the sample mean is more accurate and a reflection of the actual population mean.

Table 7. Specific indirect effects and its confidence limits

Effects from Education Level to HIV Outcome	Estimate	S.E	Est/S.E	STTD	STDY
Direct	0.224	0.077	3.355	0.1442	0.032
Total Indirect	0.640	0.066	9.767	0.0783	0.023
Specific Indirect	0.640	0.066	9.767	0.0238	0.074
Total	0.865	0.025	35.168	0.065	0.142

CONCLUSION

In this study, the methods for developing the model and the fitting of the model to KAIS data in order to determine the appropriate probability distribution for data set with the aim of establishing the strength of the association between the categorical covariate(s) and the binary response through latent continuous mediator variable(s) using the structural equation modeling is presented. The direct, indirect and total effects was examined in a single model. Specifically, the sexual behavior patterns (intermediate variable) were used to explain why an independent variable (education level) affects the outcome HIV (status). The results indicated that there is a relationship between education attainment and knowledge /awareness of HIV/AIDS. In most research findings a relationship between high level of education and awareness of HIV has been observed such that those who have attained higher levels of education happen to be just as knowledgeable in HIV in terms of transmission, prevention, infection and control: However, the analysis suggests that education levels is not associated with HIV prevalence after controlling for a number of socio-demographic characteristics and behavioral factors, unlike with other diseases higher levels of education do not appear to be protective against HIV positivity.

Conflict of interest

There is no conflict of interest

REFERENCES

1. Joint, U. (2014). Programme on HIV/AIDS. The gap report. Geneva: Joint UN Programme on HIV. AIDS, 7.

2. National AIDS and STI Control Programme. (2018). Kenya HIV Estimate Report October 2018.

3. Karim, S. S. A., Karim, Q. A., & Baxter, C. (2015). Antibodies for HIV prevention in young women. Current opinion in HIV and AIDS, 10(3), 183-189.

4. Bollen, K. A. (1989). Structural equations with latent variables (Vol. 210). John Wiley & Sons.

5. Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. Psychometrika, 69, 167-190.

6. Knott, M., & Bartholomew, D. J. (1999). Latent variable models and factor analysis (Vol. 7, No. 2nd). Edward Arnold.

7. Tabachnick B.G., and Fidell S.L. (2007) Using Multivariate Statistics 2nd ed. California State University.

8. Hayes, A. F., Preacher, K. J., & Myers, T. A. (2014). Mediation and the estimation of indirect effects in political communication research. In Sourcebook for political communication research (pp. 434-465). Routledge.

9. Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49(1), 115-132.

10. Winship, C., & Mare, R. D. (1983). Structural equations and path analysis for discrete data. American Journal of Sociology, 89(1), 54-110.

11. Krueger, R. F., Nichol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., & McGue, M. (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. Psychological Assessment, 16(2), 107 – 119.

12. Li, J. (2005). An examination of a structural equation model of readiness to complementary and alternative medicine use among Australian university students. University of Maryland, College Park.

13. MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. Prevention science, 1(4), 173-181.

14. Mackinnon D, P. et al (2008) An introduction to statistical mediation analysis. Quantitative, Qualitative, neuropsychological, and biological, pp 313-331.

15. Mackinnon, P. David (2008). Introduction to Statistical Mediational Analysis. Taylor and Francis group, LLC.

16. Conway, C., Hammen, C., & Brennan, P. (2012). A comparison of latent class, latent trait, and factor mixture models of DSM-IV borderline personality disorder criteria in a community setting: Implications for DSM-5. Journal of Personality Disorders, 26(5), 793-803.

17. Bryan, A., Schmiege, S. J., & Broaddus, M. R. (2007). Mediational analysis in HIV/AIDS research: Estimating multivariate path analytic models in a structural equation modeling framework. AIDS and Behavior, 11, 365-383.

18. West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies.