



## OPEN ACCESS

## EDITED BY

Lubna Pinky,  
Meharry Medical College, United States

## REVIEWED BY

Ana Luisa Patrão,  
University of Porto, Portugal  
Laboni Akter,  
Khulna University of Engineering &  
Technology, Bangladesh  
Haithem Taha Mohammed Ali,  
University of Zakho, Iraq

## \*CORRESPONDENCE

Tigabu Kidie Tesfie  
✉ [tigabukidie@gmail.com](mailto:tigabukidie@gmail.com)

RECEIVED 23 January 2024

ACCEPTED 20 May 2024

PUBLISHED 24 June 2024

## CITATION

Tesfie TK, Yehuala TZ, Agimas MC,  
Yismaw GA, Wubante SM, Fente BM and  
Derseh NM (2024) Predicting the  
individualized risk of human  
immunodeficiency virus infection among  
sexually active women in Ethiopia using a  
nomogram: prediction model development  
and validation.

*Front. Public Health* 12:1375270.

doi: 10.3389/fpubh.2024.1375270

## COPYRIGHT

© 2024 Tesfie, Yehuala, Agimas, Yismaw,  
Wubante, Fente and Derseh. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Predicting the individualized risk of human immunodeficiency virus infection among sexually active women in Ethiopia using a nomogram: prediction model development and validation

Tigabu Kidie Tesfie<sup>1\*</sup>, Tirualem Zeleke Yehuala<sup>2</sup>,  
Muluken Chanie Agimas<sup>1</sup>, Getaneh Awoke Yismaw<sup>1</sup>,  
Sisay Maru Wubante<sup>2</sup>, Bezawit Melak Fente<sup>3</sup> and  
Nebiyu Mekonnen Derseh<sup>1</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, <sup>2</sup>Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia, <sup>3</sup>Department of General Midwifery, School of Midwifery, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

**Introduction:** Women are more vulnerable to HIV infection due to biological and socioeconomic reasons. Developing a predictive model for these vulnerable populations to estimate individualized risk for HIV infection is relevant for targeted preventive interventions. The objective of the study was to develop and validate a risk prediction model that allows easy estimations of HIV infection risk among sexually active women in Ethiopia.

**Methods:** Data from the 2016 Ethiopian Demographic and Health Survey, which comprised 10,253 representative sexually active women, were used for model development. Variables were selected using the least absolute shrinkage and selection operator (LASSO). Variables selected by LASSO were incorporated into the multivariable mixed-effect logistic regression model. Based on the multivariable model, an easy-to-use nomogram was developed to facilitate its applicability. The performance of the nomogram was evaluated using discrimination and calibration abilities, Brier score, sensitivity, and specificity. Internal validation was carried out using the bootstrapping method.

**Results:** The model selected seven predictors of HIV infection, namely, age, education, marital status, sex of the household head, age at first sex, multiple sexual partners during their lifetime, and residence. The nomogram had a discriminatory power of 89.7% (95% CI: 88.0, 91.5) and a calibration *p*-value of 0.536. In addition, the sensitivity and specificity of the nomogram were 74.1% (95% CI: 68.4, 79.2) and 80.9% (95% CI: 80.2, 81.7), respectively. The internally validated model had a discriminatory ability of 89.4% (95% CI: 87.7, 91.1) and a calibration *p*-value of 0.195. Sensitivity and specificity after validation were 72.9% (95% CI: 67.2, 78.2) and 80.1% (95% CI: 79.3, 80.9), respectively.

**Conclusion:** A new prediction model that quantifies the individualized risk of HIV infection has been developed in the form of a nomogram and internally validated. It has very good discriminatory power and good calibration ability.

This model can facilitate the identification of sexually active women at high risk of HIV infection for targeted preventive measures.

#### KEYWORDS

HIV, sexually active women, prediction model, Ethiopia, nomogram

## Introduction

By the end of 2022, the number of people living with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) worldwide was 39 million. The World Health Organization's (WHO) African region remains most severely affected, with approximately 1 in every 25 adults living with HIV and accounting for two-thirds of the global morbidity burden (1). This makes HIV/AIDS one of the most serious worldwide public health problems (2). Sub-Saharan Africa (SSA) is home to only 12% of the global population, yet it accounts for 71% of the global burden of HIV infection (3). Women are disproportionately affected by HIV transmission, accounting for 54% of the total number of people living with HIV (4). According to the Joint United Nations Programme on HIV/AIDS (UNAIDS) global report of 2022, the southern and eastern parts of Africa remain the most heavily affected region by HIV, with 20.6 million (54% of the global burden) people living with HIV (5).

The global HIV/AIDS burden reflects the continued transmission of HIV as a public health problem despite reductions in incidence and expanded access to antiretroviral treatment, which have helped to reduce HIV-related deaths (1). Accelerated reductions of new HIV infections are needed to end AIDS as a public health threat by 2030 (5). The world's continued commitment toward HIV/AIDS response and "ending AIDS" as a public health threat by 2030 can be signaled by the 95–95–95 targets (6). This strategy calls for 95% of people living with HIV to know their status, 95% of people with HIV diagnoses to receive antiretroviral therapy (ART), and 95% of people receiving ART to have viral suppression by 2025 (7).

The burden of HIV/AIDS remains high in Ethiopia, even though a considerable scale-up of comprehensive HIV/AIDS interventions has been made so far. According to the 2016 Global Burden of Disease study, Ethiopia achieved a 77 and 0.4% reduction between 1990 and 2016 in the incidence of infection and mortality, respectively. Ethiopia achieved the Millennium Development Goals (MDGs) toward HIV/AIDS, which aimed to reduce the incidence by 50% from 2000 to 2015 (8). In 2022, the Ethiopian Public Health Institute (EPHI) report revealed that approximately 573,538 adults were living with HIV, which makes the national adult HIV prevalence of 0.91% (9). This shows the presence of high infection transmission that poses a threat to target achievements. To achieve

HIV/AIDS-related Sustainable Development Goals in Ethiopia, concerted efforts are required for comprehensive HIV/AIDS interventions (8).

HIV/AIDS is a threat to socioeconomic development beyond morbidity and mortality. It is affecting educational achievements, productivity and economic growth, social status, and other components of human development, including human rights and gender equality (10–12). Hence, HIV/AIDS demands a greater degree of well-thought-out, intensive, and coordinated action by sub-national, national, and international agencies (13, 14). HIV/AIDS in women has a strong implication for national and global socioeconomic development since women play a crucial role in development activities such as trade, agriculture, and family support (15).

According to the studies conducted in different countries across Africa, several factors were found to be significantly associated with HIV infection including age (16–18), marital status (16, 18, 19), educational attainment (17), wealth index (20), sex of the household head (21), residence (17, 18), region (17), number of sexual partners (17, 18), age at first sex (22), and sexually transmitted infections (19). However, none of these studies apply predictive modeling techniques to assess which combination of variables best predicts HIV infection risk.

Unprotected heterosexual encounters are the primary mechanism for HIV transmission in Africa. Hence, consistent condom use remains the most effective method of HIV prevention. However, according to studies conducted among Mozambican women, consistent condom use depends on sociodemographic characteristics that include educational status, marital status, and age. In addition, psychosocial factors such as perception of barriers to safe sex, condom use negotiation self-efficacy, and HIV prevention knowledge play a significant role in the utilization of condoms (23, 24). In addition, three mechanisms, namely, pre-exposure prophylaxis, behavioral change communications, and early initiation of ART remain vital strategies for preventing HIV/AIDS transmission (25).

Identification of individuals at high risk for HIV/AIDS and linking them to prevention services is essential for continued progress toward ending HIV as a public health threat (26, 27). Risk estimation tools, such as nomograms, could play a role in directing targeted preventive strategies through the quantification of individualized risk (25). Predicting the risk of HIV infection is also used as an early-warning system to notify prevention programs (28). Women are biologically more prone to HIV infection for the reason that a greater mucus area is exposed to the virus during sexual intercourse. Their higher risk is also linked to gender inequalities and the economic and social pressures of poverty, which intensify HIV infection risk (29). Developing a risk prediction model for these highly vulnerable groups of the population had a vital role in the prevention of HIV infection. Prediction models can help clinicians, public health professionals, and clients by directing the decision-making process in the choice of

Abbreviations: AOR, Adjusted odds ratio; ART, Antiretroviral therapy; AUC, Area under the curve; DHS, Demographic and Health Survey; EDHS, Ethiopian Demographic and Health Survey; EPHI, Ethiopian Public Health Institute; HIV/AIDS, Human immunodeficiency virus/acquired immunodeficiency syndrome; ICC, Intra-class correlation coefficient; LASSO, Least absolute shrinkage and selection operator; NPV, Negative predictive value; PPV, Positive predictive value; STIs, Sexually transmitted infections.

possible interventions through individualized risk stratification with the hope of improving patient outcomes and quality of care (30, 31).

Unlike previous studies conducted among sexually active women (16–21) that primarily focused on investigating and identifying statistically significant associations between factors and HIV status, we utilize a predictive modeling technique to identify a combination of factors that best predicts the individualized risk of HIV infection among sexually active women, develop user-friendly nomogram based on the best combination of variables, evaluate the extent to which the nomogram can differentiate between negative and positive HIV status, assess the agreement between predicted probability and actual probability of HIV infection, and assess the internal validity of the predictive model developed for individualized HIV infection risk estimation. Therefore, this study aims to develop and validate a nomogram for an individualized HIV infection risk estimation among sexually active women in Ethiopia using nationally representative data.

## Methods

### Study design, setting, and period

A nationwide, cross-sectional Demographic and Health Survey (DHS) was conducted in Ethiopia between 18 January and 27 June 2016 (32).

### Sampling procedure, population, and data sources

Participants in the 2016 Ethiopian Demographic and Health Survey (EDHS) were identified using a two-stage stratified cluster sampling technique. In the first stage of the sampling, 645 clusters (202 urban and 443 rural) were selected based on the 2007 Ethiopian Population and Housing Census sampling frame. In the second stage, a fixed number of 28 households per cluster were randomly selected from the household list. The source population consisted of all sexually active women in Ethiopia, whereas the study population was all sexually active women in the selected enumeration areas. A total of 15,683 women of reproductive age were interviewed. Out of these, 3,721 women had never had sex. From a total of 11,962 women who read the consent statement, 11,185 (a weighted sample of 10,253) women granted permission for HIV testing. A total weighted sample of 10,253 sexually active women were included in this study. Details of the survey methodology have been published elsewhere (32). The data used for this study were obtained from the DHS program using the link <http://www.dhsprogram.com>.

### Variables of the study

The outcome variable was HIV status (positive or negative). To confirm the diagnosis, interviewers collected blood specimens by finger pricking from sexually active women who granted consent for HIV testing. Blood samples were dried overnight, packed the following morning for storage, and then transported to the EPHI laboratory. Until further testing, blood samples were stored at  $-20^{\circ}\text{C}$ . The HIV testing algorithm used to determine HIV status has

been published elsewhere (32). The predictor variables included age, education, marital status, household wealth index, sex of the household head, residence, region, age at first sex, multiple sexual partners during the lifetime, multiple sexual partners in 1 year, diagnosed STIs, genital ulcers, and genital discharge (16–21). For all predictor variables, ascertainment was made based on women's responses to the interviewer-administered questionnaire.

Regarding categorization, predictor variables were categorized as follows: age (15–19, 20–24, 25–29, 30–34, 35–39, 40–44, and 45–49), education (no, primary, secondary, and higher), marital status (single, married/cohabiting, and separated/widowed/divorced), wealth index (poorest, poorer, middle, richer, and richest), sex of the household head (male or female), residence (urban or rural), region (Tigray, Afar, Amhara, Oromia, Somali, Benishangul-Gumuz, South Nation Nationalities and Peoples' Region (SNNPR), Gambela, Harari, Addis Ababa, and Dire Dawa), age at first sex in years ( $\leq 14$ , 15–17, and  $\geq 18$ ), multiple sexual partners during lifetime (no or yes), multiple sexual partners in 1 year (no or yes), diagnosed STIs (no or yes), genital ulcers (no or yes), and genital discharge (no or yes).

### Data processing and analysis

Data obtained from the DHS program were imported into Stata version 17 software. The individual and HIV datasets were merged using the cluster number (v001), household number (v002), and respondent's line number (v003). Samples were weighted using the weight variable (hiv05) to maintain representativeness. Missing data were managed according to the DHS guidelines (33). The results are presented in the form of text, tables, and figures.

During the model development process, variables were screened for chi-square assumptions. Hence, variables that fulfilled the chi-square assumptions were entered into the least absolute shrinkage and selection operator (LASSO) regression model. Then, variables with non-zero coefficients from the LASSO model were used in the multivariable analysis. DHS data have a clustering nature, in which women from the same cluster are more dependent than women from different clusters. We used the intra-class correlation coefficient (ICC) to estimate the clustering effect. The ICC was greater than 10%, indicating that considering the clustering effect is mandatory to obtain appropriate estimates of association for the development of the prediction model. Therefore, the multivariable mixed-effects binary logistic regression model was fitted. Insignificant variables were reduced one by one using the likelihood ratio test. Based on the predictors retained in the reduced model, an individualized HIV infection risk estimation tool was developed in the form of a nomogram.

The nomogram's performance was evaluated using a variety of metrics. The Brier score was used to indicate the overall model performance, the area under the curve (AUC) of the receiver operating characteristics for model discriminating ability assessment, and a calibration plot for detecting the agreement between the observed and predicted probability of developing HIV infection (34). Additionally, accuracy, sensitivity, specificity, and predictive values were also calculated to assess the performance of the model (35). In order to choose the optimal probability cutoff point for risk stratification as high or low, the Youden's index method was applied (36). An individualized risk stratification table was constructed based on the identified cutoff point to assign each woman to either the high- or

low-risk group and to facilitate prevention interventions based on the evaluated risk of developing the HIV infection (37). The internal validity of the model was evaluated using the bootstrapping method (38). Bootstrapping provides stable estimates with low bias as compared to other methods of internal validation, such as the split sample (39). The difference between the apparent performance (performance before validation) and true performance (performance after validation) was calculated. The transparent reporting of a multivariable prediction model for an individual prognostic or diagnostic checklist was used to report the findings of the study (40).

The overall modeling process is presented in Figure 1.

## Data collection and instruments

The 2016 EDHS data were collected through face-to-face, interviewer-administered questionnaires. Both women's and bio-marker questionnaires, which include sociodemographic and economic factors, behavioral characteristics, and HIV-testing results were used for data collection. During HIV testing, all samples were tested with the Genscreen ULTRA Ag/Ab (Bio-Rad) enzyme-linked immunoassay (ELISA I). All samples that tested positive for ELISA I were subjected to a second ELISA (ELISA II) and the Bioelisa HIV 1 + 2 Ag/Ab combination (Biokit). In total, 5% of the samples that tested negative for ELISA I were also subjected to ELISA II, while the other 95% were recorded as negative. Concordant negative results on ELISA I and ELISA II were recorded as negative. If the results of ELISA I and ELISA II were discordant, the specimen was considered inconclusive. Concordant positive results on ELISA I and ELISA II were also subjected to a third confirmatory assay. Accordingly, if an

individual had positive test results for both ELISA I and II, the final HIV test result was considered to be positive, if the confirmatory test (Inno-Lia) was positive. On the other hand, the final HIV test was taken as an inconclusive result for an individual who had positive test results for both ELISA I and II if the Inno-Lia was a negative or indeterminate result (32).

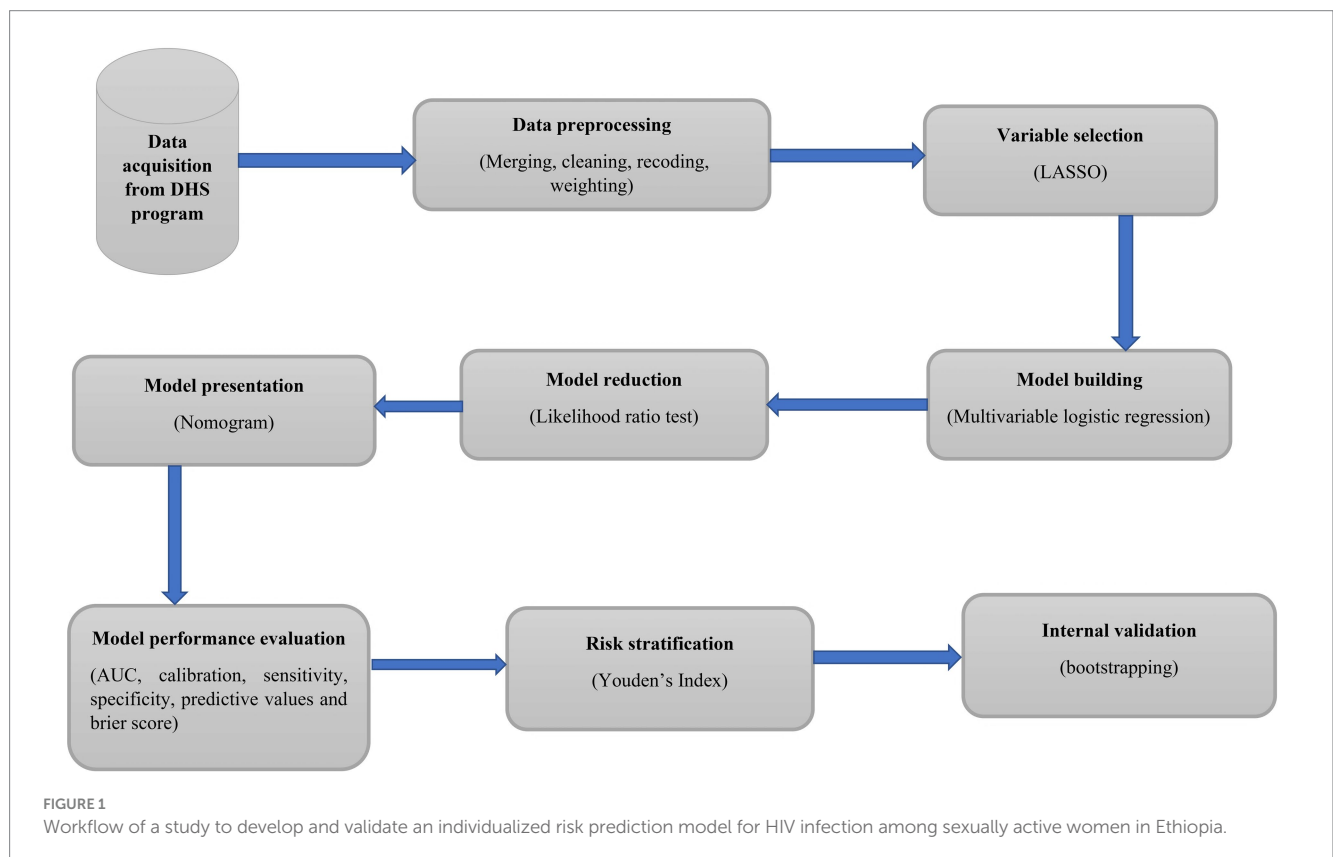
## Ethical considerations

During the 2016 EDHS survey, all participants in the HIV test granted their consent for blood sample collection. Individuals who did not grant consent for participation in the HIV test were not included (32). In this study, we used publicly available data with no personal identifiers, which was provided by the DHS program. All methods were carried out in accordance with the relevant guidelines of the DHS program.

## Results

### Characteristics of participants

The mean (SD) age of participants was 31 (8.3) years, of which more than one-fifth of women were between 25 and 29 years old. Approximately 6 out of 10 women (58.9%) did not attend formal education. The majority (84.9%) of the participants were married/cohabiting during the survey. Approximately one-fourth (23.3%) of participants lived in households with the richest wealth index. More than three-fourths (78.2%) of women lived in male-headed



households. Regarding age at first sex, 23.3% of women started sexual intercourse at or before 14 years of age. A total of 14.8 and 24.8% of sexually active women had multiple sexual partners in the last year

prior to the survey and in their lifetime, respectively. The majority (80.6%) of women resided in urban areas (Table 1).

TABLE 1 Characteristics of study participants, EDHS 2016.

Variable	Category	Weighted frequency (n = 10,253)	Percentage (%)
Age	15–19	693.6	6.8
	20–24	1741.4	17.0
	25–29	2315.0	22.6
	30–34	1946.2	19.0
	35–39	1603.6	15.6
	40–44	1085.5	10.6
	45–49	867.4	8.5
Education	No	6035.5	58.9
	Primary	3013.1	29.4
	Secondary	731.5	7.1
	Higher	472.5	4.6
Marital status	Single	338.4	3.3
	Married/cohabiting	8702.5	84.9
	Widowed/divorced/separated	1211.7	11.8
Wealth index	Poorest	1918.3	18.7
	Poorer	1974.6	19.3
	Middle	2015.7	19.7
	Richer	1960.7	19.1
	Richest	2383.3	23.3
Sex of the household head	Male	8016.9	78.2
	Female	2235.7	21.8
Age at first sex (years)	≤14	2386.9	23.3
	15–17	4436.2	43.3
	≥18	3429.5	33.4
Multiple sexual partners in the last year	Yes	1514.7	14.8
	No	8738.0	85.2
Multiple sexual partners during lifetime	Yes	2441.3	23.8
	No	7811.3	76.2
Diagnosed STIs	Yes	28.7	0.3
	No	10223.9	99.7
Genital ulcers	Yes	208.0	2.0
	No	10044.6	98.0
Genital discharge	Yes	262.1	2.6
	No	9990.5	97.4
Residence	Urban	1994.2	19.4
	Rural	8258.4	80.6

## National and regional prevalence of HIV infection

The overall prevalence of HIV among sexually active women in Ethiopia was 1.5% (95% CI; 1.3, 1.8%). The lowest and highest prevalences were found in the Somali (0.1%) and Gambela (6.7%) regions, respectively (Table 2).

## Development of an individualized HIV risk prediction model

### Predictor selection

A total of 4 out of the 13 predictor variables (STI, genital ulcer, genital discharge, and region) did not meet the chi-square assumption and were therefore not included in the LASSO regression model. The remaining nine variables were entered into LASSO regression, and non-zero coefficients were identified at a tuning parameter (lambda) of 0.0004 with a cross-validated mean deviance of 0.20 and an out-of-sample deviance ratio of 0.18 at the 46th iteration. All variables included in the LASSO had non-zero coefficients and were thus included in the multivariable model.

### Multivariable mixed-effects logistic regression model

The ICC value was estimated to be 37.2% (95% CI: 28.9, 46.4). Therefore, to account for the variability explained by the clustering effect, variables with non-zero coefficients from the LASSO regression were exported into a mixed-effects multivariable binary logistic regression model. To achieve model parsimony, further model

TABLE 2 National and regional prevalence of HIV infection among sexually active women in Ethiopia, EDHS 2016.

Region	Sampled women	Weighted HIV-positive frequency	Percentage of HIV (%)
Tigray	749.0	13.8	1.8
Afar	94.3	1.6	1.7
Amhara	2538.3	41.0	1.6
Oromia	3829.2	44.2	1.2
Somali	310.1	0.3	0.1
Benishangul Gumuz	109.5	2.1	1.9
SNNPR	2012.8	13.6	0.7
Gambela	31.4	2.1	6.7
Harari	26.0	1.1	4.2
Addis Ababa	494.5	32.2	6.5
Dire Dawa	57.5	2.6	4.5
Ethiopia (Total)	10,253	154.5	1.5

reduction was conducted by reducing insignificant variables ( $p$ -value  $>0.05$ ) using the likelihood ratio test. Two variables, namely, household wealth index ( $p$ -value = 0.076) and multiple sexual partners in 1 year ( $p$ -value = 0.153), were removed. The full model (model with nine predictors) and the reduced model (model with seven predictors) were compared using the likelihood ratio test, which implies that no statistically significant difference was observed between the two models (likelihood ratio chi-square = 6.98 and  $p$ -value = 0.222). Finally, seven predictors were retained and used for nomogram development (Table 3).

The probability of developing HIV infection among sexually active women was predicted using the following regression formulas described in Equations 1, 2.

$$\begin{aligned}
 &\text{Linear predictor of the model (lp)} = -7.30 + 1.44 \\
 &*age(30 - 34) + 1.30 * age(35 - 39) + 1.15 \\
 &*age(40 - 44) + 0.97 * age(45 - 49) + 0.92 \\
 &*education(primary) + 0.87 * education(secondary) + 0.87 \\
 &*marital\ status(separated / divorced / widowed) + 0.72 \\
 &*sex\ of\ household\ head(female) + 0.42 \\
 &*age\ at\ first\ sex(\leq 14\ years) + 0.96 \\
 &*lifetime\ multiple\ sexual\ partners(yes) + 1.46 \\
 &*residence(urban)...
 \end{aligned}
 \tag{1}$$

Consequently, the probability of being HIV<sup>+</sup> is given by the following equation:

$$HIV+ = \frac{\exp^{lp}}{1 + \exp^{lp}} \tag{2}$$

### Nomogram for predicting individualized HIV infection risk

For ease of clinical applicability, the nomogram was developed using the seven predictors from the reduced model. Scores corresponding to each predictor were generated from the nomogram division table automatically. Based on the total score, the individualized probability of developing HIV infection could be calculated easily using the nomogram (Figure 2).

### Performance of the nomogram

The role of each predictor for the overall AUC was presented (Figure 3A). The AUC of the nomogram for individualized HIV risk prediction was 89.7% (95% CI: 88.0, 91.5) (Figure 3B). When

TABLE 3 Adjusted odds ratios from multivariable mixed-effects logistic regression model, EDHS 2016.

Variables	Category	AOR (95% CI)	$\beta$ -coefficients	$p$ -value
Age group	15-19	Ref		
	20-24	0.76 (0.30, 1.94)	-0.28	0.562
	25-29	2.15 (0.93, 4.97)	0.77	0.073
	30-34	4.21 (1.85, 9.59)	1.44	0.001
	35-39	3.67 (1.60, 8.41)	1.30	0.002
	40-44	3.16 (1.34, 7.43)	1.15	0.008
	45-49	2.64 (1.08, 6.44)	0.97	0.032
Educational status	No	Ref		
	Primary	2.50 (1.77, 3.53)	0.92	< 0.001
	Secondary	2.39 (1.53, 3.73)	0.87	< 0.001
	Higher	1.21 (0.67, 2.19)	0.19	0.535
Marital status	Single	Ref		
	Married/cohabiting	1.21 (0.67, 2.18)	0.19	0.525
	Widowed/divorced/separated	2.39 (1.33, 4.32)	0.87	0.004
Sex of the household head	Male	Ref		
	Female	2.05 (1.48, 2.85)	0.72	< 0.001
Age at first sex	$\leq 14$	1.52 (1.05, 2.21)	0.42	0.026
	15-17	1.21 (0.89, 1.68)	0.20	0.224
	$\geq 18$	Ref		
Multiple sexual partners during lifetime	Yes	2.61 (1.98, 3.45)	0.96	< 0.001
	No	Ref		
Residence	Urban	4.29 (2.97, 6.20)	1.46	< 0.001
	Rural	Ref		
Intercept		0.0006 (0.0002, 0.0018)	-7.30	< 0.001

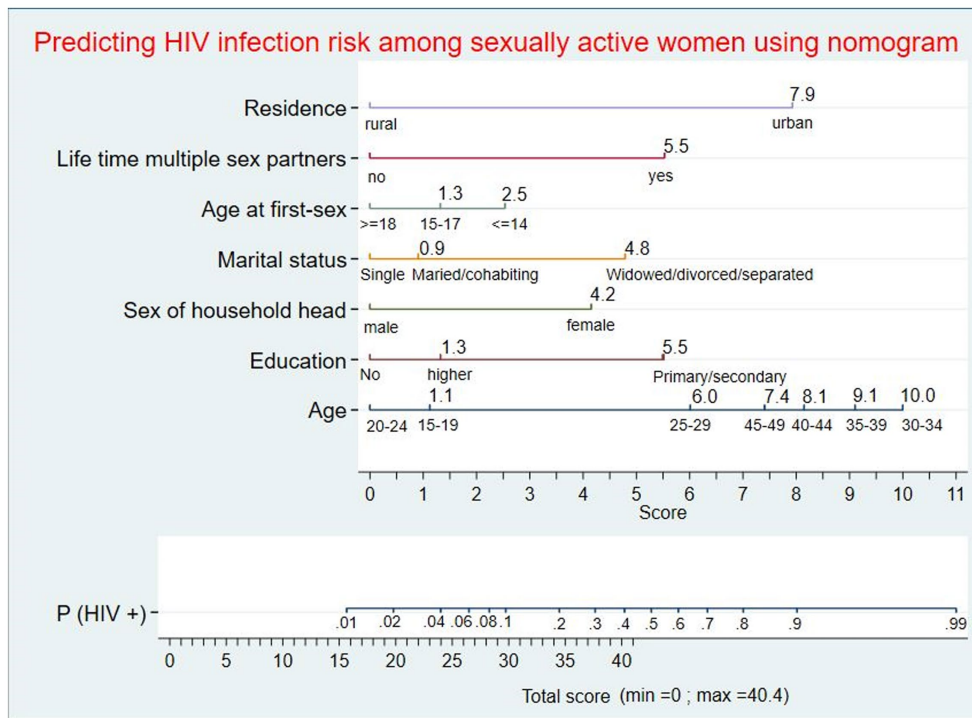


FIGURE 2

Nomogram for predicting individualized risk of HIV infection among sexually active women in Ethiopia. An individual with a total score of 21 will correspond to a 0.0281 probability of developing HIV infection; women with a  $\geq 0.0281$  probability will be classified as high-risk, whereas women with a  $< 0.0281$  probability will be classified as low-risk during nomogram application.

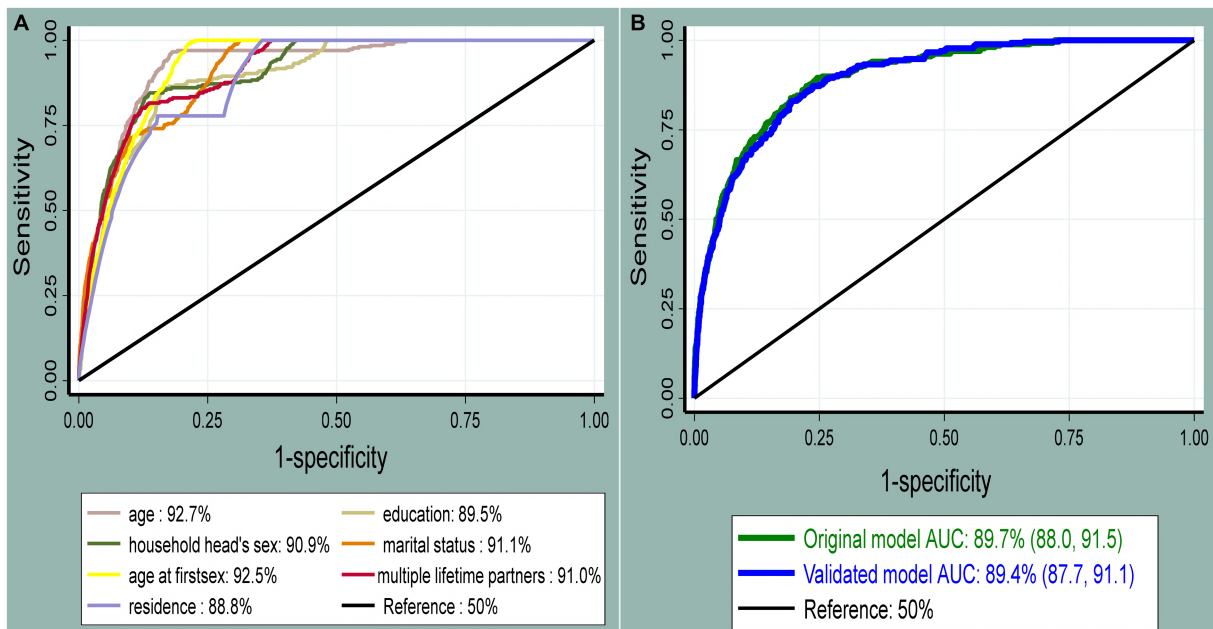
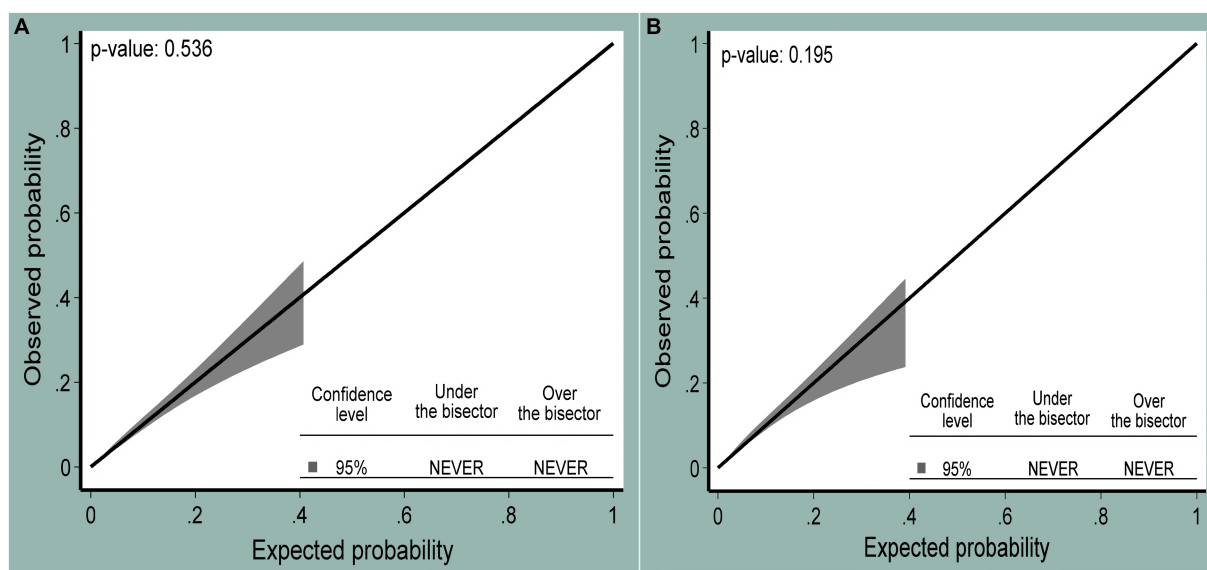


FIGURE 3

The receiver operating characteristic curve shows the performance of each predictor (A) and the original and internally validated model (B). The diagonal black line represents a model that discriminates by chance (AUC = 50%); the x-axis shows the proportion of individuals without HIV infection who were incorrectly classified as having HIV (false-positive rate), and the y-axis shows the proportion of individuals with HIV who were correctly classified as having HIV infection (true-positive rate).



**FIGURE 4**  
The calibration plot showing the agreement between observed (y-axis) and expected (x-axis) probabilities of being HIV positive at 95% confidence levels for the original (A) and internally validated models (B).

additional performance measures were assessed, the model had a sensitivity of 74.1% (95% CI: 68.4, 79.2), specificity of 80.9% (95% CI: 80.2, 81.7), PPV of 8.7% (95% CI: 7.6, 9.9), NPV of 99.2% (95% CI: 99.0, 99.4), and an accuracy of 80.8% (95% CI: 80.0, 81.5) to identify individuals at risk for HIV at the 0.0281 probability cutoff point identified by Youden’s index (max J=0.55). The calibration test had a p-value of 0.536, indicating no significant difference between the observed probability of developing HIV and the expected probability of developing HIV (the calibration plot includes the bisector) (Figure 4A). The Brier score of the model was 0.02.

### Individualized HIV infection risk stratification table

The risk stratification table was constructed using Youden’s index (max J=0.55), and the corresponding probability cutoff point with this index was 0.0281. The risk was dichotomized into low- (<0.0281) and high-risk ( $\geq 0.0281$ ) groups. Of the total participants, 1,249 (12.2%) were in the high-risk group. Of all HIV infections, the majority (73.4%) of HIV-positive individuals were from the high-risk category. The prevalence of HIV was 0.5 and 9.1% in low-risk and high-risk groups, respectively (Table 4).

### Internal validation

The developed nomogram was internally validated by bootstrapping using 1,000 bootstrap samples with replacement to determine the degree of overfitting (i.e., models performing better in the development sample than in the new sample after bootstrapping). The AUC of the internally validated model was 89.4% (95% CI: 87.7, 91.1) (Figure 3B). Additional model performances were also evaluated after internal validation. The model’s sensitivity was 72.9% (95% CI: 67.2, 78.2), the specificity was 80.1% (95% CI: 79.3, 80.9),

**TABLE 4** Risk stratification table based on the probability of HIV infection identified by the nomogram.

Risk category (probability)	Nomogram	
	Frequency	Frequency (prevalence) of HIV <sup>+</sup>
Low risk ( $p < 0.0281$ )	9,004 (87.8%)	41 (0.5%)
High risk ( $p \geq 0.0281$ )	1,249 (12.2%)	113 (9.1%)
Total	10,253 (100%)	154 (1.5%)

the PPV was 8.2% (95% CI: 7.1, 9.4), the NPV was 99.2% (95% CI: 99.0, 99.4), and the accuracy was 80.0% (95% CI: 79.2, 80.7) at the 0.0281 probability cutoff point identified by a Youden’s index of 0.53. The calibration curve p-value was 0.195, which indicated the presence of agreement between the observed and predicted probability of mortality across all probability thresholds (bisector) (Figure 4B). The optimism coefficient and Brier score were 0.003 and 0.02, respectively.

### Discussion

This study aimed to develop an individualized risk prediction tool that helps predict the probability of developing HIV infection for a sexually active woman. Sexually active women are a vulnerable population group for HIV. Women are more prone to HIV infection than men due to biological differences and gender inequalities (29). Developing a risk prediction model for this highly vulnerable group of the population plays a vital role in the prevention of HIV infection and subsequent reduction of HIV prevalence. This will make a great contribution to the global target of ending HIV as a public health threat. Prediction models direct the decision-making process in the choice of interventions through the estimation of individualized risk with the aim of achieving better outcomes (30, 31, 41).



The nomogram risk prediction model was constructed using multivariable logistic regression analysis, considering the clustering effect. Seven predictors were selected for model development after passing through LASSO regression and model reduction. The developed model was assessed for internal validity using a bootstrapping technique. The developed nomogram could assist health professionals in their delivery of HIV prevention services. Nomograms are easy-to-use graphical tools during the application (42).

The odds of having HIV infection among sexually active women were higher among older age groups compared with women aged 15–19 years. This finding is supported by studies from Nigeria (16) and Kenya (18). The positive relationship between increased age and HIV infection might be because older women have relatively more repeated exposure to risk factors for HIV than younger women. In Ethiopia, higher HIV prevalence is reported in older age groups (43).

Those with primary and secondary education had higher odds of HIV infection as compared to women who had no education. Previous studies have found that better levels of education are related to an increased risk of HIV infection due to the fact that more educated women are more likely to be mobile and have more sexual partner networks (17). Being widowed, divorced, or separated was found to be significantly associated with higher odds of HIV infection. This finding is supported by studies from Nigeria (16) and Zambia (44). Based on a recently conducted population survey in the SSA, widowed and divorced women are more likely to engage in high-risk behaviors for substance abuse which may increase the probability of engaging in risky sexual behaviors (45). The sex of the household head was also significantly associated with HIV infection. Sexually active women living in female-headed households have higher odds of HIV infection as compared to sexually active women living in male-headed households. This finding is supported by previous studies from Ethiopia (21) and SSA (46). In Ethiopia, men are usually heads of household. Nevertheless, when women are heads of household, it is a sign that the woman's marital status is unmarried, divorced, or widowed. The probability of women engaging in adopting high-risk sexual activities, including multiple partners, engagement, in paid sex, and sexual harassment. Women who started having sex at or before the age of 14 were more likely to contract HIV. This finding is supported by evidence from Tanzania (22), Zimbabwe (47), and Rwanda (48). This may be because those who start having sex early are more likely to have multiple sexual partners in their lifetime. Having multiple sexual partners in life increases the odds of HIV infection. This is consistent with findings from other studies (17, 18). Sexually active women residing in urban areas had higher odds of HIV infection compared to sexually active women living in rural areas. This is supported by another study from Ethiopia (49). In Ethiopia, urban areas had a higher prevalence of HIV infection, which may be a possible reason for the higher transmission of HIV among these residents (43).

The optimal combination of variables to predict HIV infection using a nomogram was age group, education, marital status, sex of the household head, age at first sex, multiple sexual partners during the lifetime, and place of residence. The combined performance of these women's sociodemographic characteristics and sexual behavior resulted in a prediction nomogram with an AUC of 89.7% (95% CI: 88.0, 91.5), which was well-calibrated ( $p$ -value = 0.536). This HIV risk prediction model had very good discriminatory power,

according to the prediction model performance classification (50). The performance of this nomogram was better than that of a study conducted in Chicago that developed a predictive model for identifying women at risk for HIV, which had an AUC of 74% (95%CI: 67, 81). Variables included were as follows: STIs, substance use, hepatitis C, pregnancy, race, ethnicity, age group, healthcare site, and number of medical encounters (51). In addition, the performance of the developed nomogram in this study was better than a prediction risk score for HIV among adolescent girls and young women in South Africa that was conducted to identify those in need of HIV pre-exposure prophylaxis based on their elevated risk, which had an AUC of 78% in the development and 76% in the validation data, respectively (52). Variables included in a study conducted in South Africa were age, age at first sex, relationship status, education, socioeconomic status, number of sexual partners, any STI, HIV-positive partner, condom use at last sex, ever engaged in transactional sex, parent dead, ever pregnant, and ever raped (52). The variation in model performance between the two studies might be due to differences in the strength of the included predictors for model development (53).

Based on Youden's index, an optimal probability cutoff point of 0.0281 was identified and resulted in better model performance (including discrimination, sensitivity, specificity, and predictive values) compared to other cutoff probability points. Using this probability, the cutoff point for HIV risk prediction the model performance would have good benefits in a risk-stratified intervention. This means that women with a probability of at least 0.0281 should be screened and linked to HIV prevention services. Intentional efforts are required to recruit such high-risk sexually active women for prevention services to reduce their likelihood of serostatus conversion. However, the optimal probability cutoff point identified through Youden's index does not replace routine HIV screening services offered in clinical care settings. The HIV risk prediction nomogram is meant to support healthcare professionals in stratifying sexually active women based on their individualized risk for HIV infection and providing services according to their risk strata. These findings inform users of the developed nomogram that HIV services for sexually active women should particularly target those with elevated risk to ensure HIV prevention interventions are cost-effective and impactful (52, 54).

The nomogram was validated internally using the bootstrap resampling technique. The discriminatory performance of the validation sample was 89.4% (95% CI: 87.7, 91.1), and regarding calibration, it was well calibrated ( $p$ -value = 0.195). The optimism coefficient was 0.003, indicating that the nomogram is less likely to be sample-dependent. Hence, the nomogram can be used for risk prediction among sexually active women in Ethiopia.

## Policy implications and applications

Our nomogram has policy implications and applications for HIV/AIDS prevention. First, the developed nomogram may be used by clinicians and public health experts to enhance counseling and health education for sexually active women in highly HIV-prevalent areas of Ethiopia, which is an important strategy to improve HIV prevention services, including screening and pre-exposure prophylaxis. Second, a nomogram may also be useful to monitor changes in risk over time

if the sexually active woman's risk is changing from low to high or high to low. Third, this nomogram may be used to allocate more resources to areas that have a large number of high-risk sexually active women; hence, equitable resource allocation will occur.

## Strengths and limitations of the study

The strengths of this study are as follows: first, we use nationally representative data, which maintains its generalizability for all sexually active women in Ethiopia. Second, to our knowledge, this is the first nomogram developed for the prediction of HIV in the country. Third, the small optimism coefficient identified in the internal validation process indicates a less likely overfitting of the nomogram, and hence it can predict HIV when applied to an independent set of samples with very good performance. However, the limitations of this study are as follows: first, limited applicability in other countries due to a lack of external validation, thus, nomogram applicability is subjected to external validation for other countries. Second, the use of prevalence data instead of incidence data to develop a risk prediction model for sexually active women may affect the nomogram's performance in predicting new HIV infections.

## Conclusions and recommendations

Our nomogram has shown very good discrimination and good calibration to identify sexually active women with a high risk of developing HIV infection. This model directs high-risk women to be targeted for intensive HIV prevention interventions at both health facilities and community levels. Therefore, we recommend the use of this nomogram by health professionals to guide their decision-making process when providing HIV prevention services. Researchers should conduct a feasibility study on the use of the nomogram in clinical and community settings to assess its user-friendliness and accuracy in identifying HIV in high-risk sexually active women. External validation should be conducted before the application of the model in other HIV-epidemic countries.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

Ethical considerations during EDHS 2016 data collection were approved by the ICF Institutional Review Board (IRB). Moreover, EDHS protocols were reviewed by the ICF IRB and by an IRB in Ethiopia during the data collection period. Before each interview or HIV test was conducted, an informed consent statement was read to the respondent, who might accept or decline to participate. The informed consent maintained voluntary participation and autonomy while strictly maintaining confidentiality and privacy.

Individuals who did not grant consent for participation in HIV testing were not included (30). In this study, we used a publicly available data provided by the DHS program, with no personal identifiers. All methods were carried out in accordance with the relevant guidelines of the DHS program after we received a letter of permission from the Institutional Review Board of the DHS program. The studies were conducted in accordance with local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with national legislation and institutional requirements.

## Author contributions

TT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. TY: Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. MA: Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – review & editing. GY: Methodology, Software, Validation, Visualization, Writing – original draft. SW: Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. BF: Formal analysis, Software, Validation, Writing – original draft, Writing – review & editing. ND: Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

We would like to acknowledge the Demographic and Health Surveys (DHS) program for providing the dataset.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- World Health Organization (WHO). *The global health observatory: Estimated number of people (all ages) living with HIV*. Geneva: World Health Organization (2023).
- Dong Y, Liu S, Xia D, Xu C, Yu X, Chen H, et al. Prediction model for the risk of HIV infection among MSM in China: validation and stability. *Int J Environ Res Public Health*. (2022) 19:1010. doi: 10.3390/ijerph19021010
- Kharsany AB, Karim QA. HIV infection and AIDS in sub-Saharan Africa: current status, challenges and opportunities. *Open AIDS J*. (2016) 10:34–48. doi: 10.2174/1874613601610010034
- The joint United Nations Programme on HIV/AIDS (UNAIDS). *Global HIV and AIDS statistics fact sheet*. UNAIDS (2023).
- Joint United Nations Programme on HIV/AIDS. *In Danger: UNAIDS global AIDS update 2022*. Geneva: UNAIDS (2022).
- Frescura L, Godfrey-Faussett P, Feizzadeh AA, El-Sadr W, Syarif O, Ghys PD. Achieving the 95 95 95 targets for all: a pathway to ending AIDS. *PLoS One*. (2022) 17:e0272405. doi: 10.1371/journal.pone.0272405
- Heath K, Levi J, Hill A. The joint United Nations Programme on HIV/AIDS 95–95–95 targets: worldwide clinical and cost benefits of generic manufacture. *AIDS*. (2021) 35:S197–203. doi: 10.1097/QAD.0000000000002983
- Deribew A, Biadgilign S, Deribe K, Dejene T, Tessema GA, Melaku YA, et al. The burden of HIV/AIDS in Ethiopia from 1990 to 2016: evidence from the global burden of diseases 2016 study. *Ethiop J Health Sci*. (2019) 29:859–68. doi: 10.4314/ejhs.v29i1.7
- The Ethiopian Public Health Institute. *HIV related estimates and projections in Ethiopia for the year 2021–2022* EPHI Ethiopian Public Health Institute (EPHI). (2022).
- Boutayeb A. The impact of HIV/AIDS on human development in African countries. *BMC Public Health*. (2009) 9:S3. doi: 10.1186/1471-2458-9-S1-S3
- Taraphdar P, Guha RT, Haldar D, Chatterjee A, Dasgupta A, Saha B, et al. Socioeconomic consequences of HIV/AIDS in the family system. *Nigerian Med J*. (2011) 52:250–3. doi: 10.4103/0300-1652.93798
- Gebre A, Yirgu D, Kloos H. Socioeconomic and psychosocial impacts of HIV/AIDS and responses at different levels of society In: G Tadele and H Kloos, editors. *Vulnerabilities, impacts, and responses to HIV/AIDS in sub-Saharan Africa*. London: Palgrave Macmillan UK (2013). 107–26.
- Spicer N, Aleshkina J, Biesma R, Brugha R, Caceres C, Chilundo B, et al. National and subnational HIV/AIDS coordination: are global health initiatives closing the gap between intent and practice? *Glob Health*. (2010) 6:3. doi: 10.1186/1744-8603-6-3
- Hushie M, Omenyo CN, van den Berg JJ, Lally MA. State-civil society partnerships for HIV/AIDS treatment and prevention in Ghana: exploring factors associated with successes and challenges. *BMC Health Serv Res*. (2016) 16:332. doi: 10.1186/s12913-016-1598-9
- Ankrah EM, Henry K. Empowering women may help retard HIV. *Network*. (1994) 15:20–1.
- Adejumo OA, Bowale A, Adesola S, Adepoju VA, Shogbamimu Y, Dacosta A, et al. Factors associated with HIV infection among clients accessing HIV counseling and testing services in a secondary referral hospital in LAGOS. *Nigeria Afr J Infect Dis*. (2020) 14:16–23. doi: 10.21010/ajid.v14i1.3
- Mulugeta SS, Wassihun SG. Prevalence of HIV/AIDS infection among sexually active women in Ethiopia: further analysis of 2016 EDHS. *AIDS Res Treat*. (2022) 2022:8971654. doi: 10.1155/2022/8971654
- Kimanga DO, Ogola S, Umuro M, Ng'ang'a A, Kimondo L, Murithi P, et al. Prevalence and incidence of HIV infection, trends, and risk factors among persons aged 15–64 years in Kenya: results from a nationally representative study. *J Acquir Immune Defic Syndr*. (2014) 66:S13–26. doi: 10.1097/QAI.0000000000000124
- Chimoyi LA, Musenge E. Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. *BMC Public Health*. (2014) 14:555. doi: 10.1186/1471-2458-14-555
- Igulot P, Magadi MA. Socioeconomic status and vulnerability to HIV infection in Uganda: evidence from multilevel modelling of AIDS Indicator survey data. *AIDS Res Treat*. (2018) 2018:1–15. doi: 10.1155/2018/7812146
- Hailu BA, Tadese F, Bogale GG, Molla A, Miheretu BA, Beyene J. Spatial patterns and associated factors of HIV Seropositivity among adults in Ethiopia from EDHS 2016: a spatial and multilevel analysis. *BMC Infect Dis*. (2020) 20:751. doi: 10.1186/s12879-020-05456-y
- Ghebremichael M, Larsen U, Paintsil E. Association of age at first sex with HIV-1, HSV-2, and other sexual transmitted infections among women in northern Tanzania. *Sex Transm Dis*. (2009) 36:570–6. doi: 10.1097/OLQ.0b013e3181a866b8
- Patrão AL, McIntyre T. Socio-demographic, marital and psychosocial predictors of safe sex behaviour among Mozambican women at risk for HIV/AIDS. *Afr J AIDS Res*. (2018) 17:323–31. doi: 10.2989/16085906.2018.1536672
- Patrão AL, McIntyre TM. Socio-demographic, marital, and psychosocial factors associated with condom use negotiation self-efficacy among Mozambican women at risk for HIV infection. *Int J Behav Med*. (2017) 24:846–55. doi: 10.1007/s12529-017-9681-0
- Fiegggen J, Smith E, Arora L, Segal B. The role of machine learning in HIV risk prediction. *Front Reprod Health*. (2022) 4:1062387. doi: 10.3389/frph.2022.1062387
- The Joint United Nations Programme on HIV/AIDS (UNAIDS). *Global AIDS strategy 2021–2026 end inequalities*. End AIDS. UNAIDS (2021).
- Jia KM, Eilerts H, Edun O, Lam K, Howes A, Thomas ML, et al. Risk scores for predicting HIV incidence among adult heterosexual populations in sub-Saharan Africa: a systematic review and meta-analysis. *J Int AIDS Soc*. (2022) 25:e25861. doi: 10.1002/jia2.25861
- Gao Y, Cai G-Y, Fang W, Li H-Y, Wang S-Y, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Commun*. (2020) 11:5033. doi: 10.1038/s41467-020-18684-2
- Mahathir M. Women at greater risk of HIV infection. *Arrows Change*. (1997) 3:1–2.
- Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis*. (2019) 11:S574–84. doi: 10.21037/jtd.2019.01.25
- Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Prog Res*. (2018) 2:11. doi: 10.1186/s41512-018-0033-6
- Central Statistical Agency (CSA) [Ethiopia] and ICF. *Ethiopia demographic and health survey 2016*. Addis Ababa: CSA and ICF (2016).
- Croft TN, Marshall AMJ, Allen CK. *Guide to DHS statistics*. Rockville, Maryland, USA: ICF (2018).
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. (2010) 21:128–38. doi: 10.1097/EDE.0b013e3181c30fb2
- Monaghan TF, Rahman SN, Agudelo CW, Wein AJ, Lazar JM, Everaert K, et al. Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value. *Medicina*. (2021) 57:503. doi: 10.3390/medicina57050503
- Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J*. (2005) 47:458–72. doi: 10.1002/bimj.200410135
- Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med*. (2008) 149:751–60. doi: 10.7326/0003-4819-149-10-200811180-00009
- Fernandez-Felix B, García-Esquinas E, Muriel A, Royuela A, Zamora J. Bootstrap internal validation command for predictive logistic regression models. *Stat J*. (2021) 21:498–509. doi: 10.1177/1536867X211025836
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. (2001) 54:774–81. doi: 10.1016/S0895-4356(01)00341-9
- Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. (2015) 162:W1–W73. doi: 10.7326/M14-0698
- Nisa SU, Mahmood A, Ujager FS, Malik M. HIV/AIDS predictive model using random forest based on socio-demographical, biological and behavioral data. *Egypt Inform J*. (2023) 24:107–15. doi: 10.1016/j.eij.2022.12.005
- Meng C, Song J, Long W, Mu Z, Sun Y, Liang J, et al. A user-friendly nomogram for predicting radioiodine refractory differentiated thyroid cancer. *Front Endocrinol*. (2023) 14:1109439. doi: 10.3389/fendo.2023.1109439
- Central Statistical Agency (CSA) [Ethiopia] and ICF. (2016) *Ethiopia demographic and health survey 2016*. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF.
- Patel CJ, Bhattacharya J, Ioannidis JPA, Bendavid E. Systematic identification of correlates of HIV infection: an X-wide association study. *AIDS*. (2018) 32:933–43. doi: 10.1097/QAD.0000000000001767
- Tenkorang EY. Marriage, widowhood, divorce and HIV risks among women in sub-Saharan Africa. *Int Health*. (2014) 6:46–53. doi: 10.1093/inthealth/ihu003
- Magadi MA. Understanding the gender disparity in HIV infection across countries in sub-Saharan Africa: evidence from the demographic and health surveys. *Sociol Health Illn*. (2011) 33:522–39. doi: 10.1111/j.1467-9566.2010.01304.x
- Pettifor AE, van der Straten A, Dunbar MS, Shiboski SC, Padian NS. Early age of first sex: a risk factor for HIV infection among women in Zimbabwe. *AIDS*. (2004) 18:1435–42. doi: 10.1097/01.aids.0000131338.61042.b8

48. Bulterys M, Chao A, Habimana P, Dushimimana A, Nawrocki P, Saah A. Incident HIV-1 infection in a cohort of young women in Butare, Rwanda. *AIDS*. (1994) 8:1585–91. doi: 10.1097/00002030-199411000-00010
49. Yilema SA, Shiferaw YA, Belay AT, Belay DB. Mapping the spatial disparities of HIV prevalence in Ethiopian zones using the generalized additive model. *Sci Rep*. (2024) 14:6215. doi: 10.1038/s41598-024-55850-8
50. Šimundić AM. Measures of diagnostic accuracy: basic definitions. *Ejifcc*. (2009) 19:203–11.
51. Friedman EE, Shankaran S, Devlin SA, Kishen EB, Mason JA, Sha BE, et al. Development of a predictive model for identifying women vulnerable to HIV in Chicago. *BMC Womens Health*. (2023) 23:313. doi: 10.1186/s12905-023-02460-7
52. Moyo RC, Govindasamy D, Manda SO, Nyasulu PS. A prediction risk score for HIV among adolescent girls and young women in South Africa: identifying those in need of HIV pre-exposure prophylaxis. *HIV Res Clin Pract*. (2023) 24:2221377.
53. Anley DT, Akalu TY, Merid MW, Tsegaye T. Development and validation of a nomogram for the prediction of unfavorable treatment outcome among multi-drug resistant tuberculosis patients in north West Ethiopia: an application of prediction modelling. *Infect Drug Resist*. (2022) 15:3887–904. doi: 10.2147/IDR.S372351
54. Takura T, Hirano Goto K, Honda A. Development of a predictive model for integrated medical and long-term care resource consumption based on health behaviour: application of healthcare big data of patients with circulatory diseases. *BMC Med*. (2021) 19:15. doi: 10.1186/s12916-020-01874-6