Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Prediction Risk Score for HIV among Adolescent Girls and Young women in South Africa: Identifying those in need of HIV Pre-Exposure Prophylaxis

Reuben Christopher Moyo (reuben.moyo2014@yahoo.com)
Stellenbosch University
Darshini Govindasamy
South African Medical Research Council
Samuel Manda
University of Pretoria
Peter Suwilakwenda Nyasulu
Stellenbosch University

Research Article

Keywords: HIV, Risk Score, Adolescent Girls and Young Women & Pre-Exposure Prophylaxis

Posted Date: October 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2069075/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background

In sub-Saharan Africa, adolescent girls, and young women (AGYW) are among those at the highest risk of acquiring HIV. Risk factors for HIV in AGYM are well studied and known in the literature. However, there is need to combine these factors into a single summary measure that could be used in the identification of the AGYW who are more likely to acquire HIV. This study aimed at developing and validating an HIV risk prediction tool for AGYW.

Methods

We analyzed existing HIV-related data on 4,399 AGYW from South Africa. The HIV risk scores were computed from summing predictor coefficients of the resulting logistic regression model. The performance of the final model at discriminating between HIV infected and non-HIV infected AGYM was assessed using area under the receiver-operating curve (AUC) and measures of discriminative abilities such as predictive values, sensitivity, and specificity. The optimal cut-point of the risk score was determined using youden index.

Results

The weighted HIV prevalence was estimated at 12.4% (11.7–14.0). Our risk scores ranged from − 1.26 to 3.80 with a mean score of 1.38 and a standard deviation of 0.86. The optimal cut-point was estimated at 1.80 with sensitivity of 62% and specificity of 70%. The prediction model's sensitivity was 15.19% and specificity of 98.92%. The model's positive predictive value was 67.42% while the negative predictive value was 88.79%. Our model performed well at predicting HIV positivity with training AUC of 0.770 and a testing AUC of 0.751.

Conclusion

Our risk score tool has shown good discrimination and calibration at predicting undiagnosed HIV. This tool could provide a simple and low-cost strategy for screening AGYW in primary health care clinics or community-based settings. The risk assessment tool will help service providers identify and link AGYW to PreP services.

Background

Adolescents and young people represent a growing share of people living with HIV worldwide. In 2019, it was estimated that over 1.7 million adolescents were living with HIV worldwide, with approximately 88% of all HIV infected adolescents living in sub-Saharan Africa (1). The World Health Organization (WHO)

estimated that 30% of all new HIV-infections globally are predicted to occur among young people aged 10–24. Adolescents aged 10–19 years account for approximately 5% of all people living with HIV worldwide, but 10% of all new HIV infections are occurring amongst these adolescents (2). In South Africa, the 2016 Demographic and Health Survey (SADHS) estimated that the HIV prevalence among the general population was 19% with wide variations at district and provincial level. HIV prevalence for KwaZulu-Natal was estimated at 29.6% while that of Limpopo was 10.4% (3). Results from the 2016 SADHS indicated that AGYW are disproportionately affected by HIV compared to their male counterparts. HIV prevalence approximately 4 times higher in AGYW compared to their male counterparts (12% versus 3% respectively).

HIV infection among AGYW is predicted by many factors broadly categorised into structural determinants and sexual behavior factors (4). Structural drivers in the context of HIV are those factors that relate to socio-economic, educational, service provision and organizational factors (5). Sexual behavior factors that predict HIV are those that relate to early sexual debut, multiple and concurrent partnership's, transactional sex and low condom use (6–10). Sexual behavior factors such as history of anal sex, having a partner suspected of having or known to be HIV and having concurrent partners were strongly correlated with HIV infections (11). Some studies have shown protective effect of parental education on HIV acquisition (8). Gender inequalities, violence against women (VAW), stigma and discrimination and limited access to sexual and reproductive health information and services are some of the structural factors that hinder adolescent girls and young women's ability to protect themselves from HIV (4). HIV infected AGYW have been shown to experience low school attendance and are associated with high school dropouts due to HIV-related morbidity (12). While HIV infection has not been shown to affect enrolment in schools, a study on HIV and educational attainments in South Africa found that Adolescent HIV infection significantly reduced their school progress index by 8.41 units (13). Among women aged 15–44, HIV one of the reading causes of death globally with higher death rates among AGYW (14).

The South Africa's National HIV, TB and STIs strategic plan prioritizes the provision of a comprehensive package of high impact, context tailored and carefully targeted combination prevention interventions (15). Combination prevention focuses on combined delivery of structural, biomedical and behavior interventions to maximize the impact of interventions on HIV incidence. Combination HIV prevention has been shown to be a key strategy in achieving the United Nations AIDS 95, 95, 95 target set in 2014. The target states that by 2020, 95 of people living with HIV will know their status, 95% of those diagnosed with HIV will receive sustained antiretroviral therapy and 95% of those receiving antiretroviral therapy will have their viral load suppressed (16). Most HIV prevention strategies mainly that focus on correct and consistent condom use and are largely male dominated which leaves women with less control (17). Evidence from studies on use of HIV Pre-Exposure Prophylaxis (PreP) for HIV prevention have shown that PreP is an effective additional preventive measure for AGYW (17, 18).

In December 2015, South Africa became the first sub-Saharan country to start implementing PreP as a biomedical HIV prevention strategy. As of 2021 it is estimated that between 160,000 to 165,000 people are using HIV PreP in ongoing and planned projects across South Africa against the target of 250,000

(19). The huge disparity between planned and achieved shows that some of the people who are supposed to be put on PreP are being missed.

Risk assessment tools have been recommended to make PreP cost effective and impactful (20). In this study, we aim to develop and validate an HIV risk assessment tool that can predict new HIV infections amongst AGYW who should be targeted with HIV PreP services.

Methods

Study Design, Setting and Population

We analyzed data on 4,399 AGYW from six South African provinces namely City of Cape Town, Ehlanzeni, OR Tambo, Tshwane, uThungulu and Zululand. The data used came from the HERStory survey conducted by South African Medical Research Council (SAMRC). These AGYW were selected to participate in the study using stratified probability proportional to size (PPS) sampling design. Sampling frames for the survey were compiled for each district based on the 2011 census small area layers and were limited to the areas targeted for the planned HIV prevention program in AYGW (21).

Outcome And Exposures Of Interest:

Outcome of Interest

HIV status: the HIV status of the AGYW ascertained using HIV testing of dried blood spots.

Predictors of the Outcome:

- Age: This was the exact age of the participant.
- Age at first Sex: This variable described the age at which the participant had her first penetrative sex.
- Condom use: This variable described whether the participant used condom at first sex, last time she had sex and whether she extensively uses condoms with her partner.
- History of Sexually transmitted infections (STIs): This variable indicated whether the participant was ever diagnosed with STIs such Syphilis and Gonorrhea.
- Number of Sexual Partners: This variable indicated the participant's number of penetrative sexual partners.
- Partner HIV status: This variable described the HIV status of the participant's sexual partner.
- Marital status: Whether the AGYW was legally or traditionally married or not.
- Transactional sex: This variable describes whether the participant engaged in sex for money or other items.
- Orphan hood: Whether the AGYW was an orphan or not.

- Use of drugs and substances: This variable described whether the participant used drugs and other substances.
- Partner age: This variable referred to the age of the participant
- Wealth index: This variable referred to the wealth index score of the AGYW
- Highest Education: This variable described the highest education level attained.
- Province: The is the exact geographical province where the AGYW was leaves.

Statistical Analysis

Data were analyzed using Stata Corp, 2019. Stata Statistical Software: Release 16. We applied sampling weights to facilitate survey data analysis. We conducted descriptive analysis using the frequency procedure to show descriptive statistics in the form of numbers and proportions. To examine association between HIV status and predictor variables, Pearson's chi-square tests (X^2) were conducted. Predictor variables were considered significant at $p < 0.05$. Missing data was not a concern because the data had only one variable with a missing observation.

Development and Validation of a Risk Prediction Model

We used 70% of the data for training the model while 30% was used for testing the performance of the model at predicting the outcome when applied to an external population. To quantify the amount of risk of HIV infection associated with each explanatory variable after controlling for the independent effects of other covariates such as age, education levels and age at sexual debut, a multivariable binary logistic regression model of HIV status on the predictors was used to obtain coefficients for use in deriving the HIV risk score. Variables in a model were selected using least absolute shrinkage selection operator (LASSO). LASSO is a feature selection method to maximize prediction performance of the model. Age, number of sexual partners and transactional sex were added in the model because they are a priori confounders and have been shown to mediate HIV infection among AGYW (22, 23). We used likelihood ratio test to select the best preforming model among models developed. The performance of the final model was assessed using discrimination and calibration methods. We used AUC to assess the performance of the model at discriminating HIV positive versus HIV negative status using testing dataset. Calibration was assessed using Hosmer Lemeshow, Brier score and Pseudo R^2. Performance of the prediction model was assessed on the testing dataset.

Scoring of a Risk Prediction Tool

Risk prediction scores were developed using beta coefficients corresponding to each predictor variable from the multivariable logit. The beta coefficients were summed to create an overall risk score. The coefficients were rounded up to two decimal points to compute a risk score (24). The optimal cut-point for the risk score and predicted probability at which AGYW is likely to have a positive HIV status was determined using youden index supported by sensitivity, specificity and AUC (25).

Results

Distribution of Study Participants

We analyzed data on 4,399 AGYW from South Africa. Overall, most of the participants came from Zululand, Ehlanzeni, O. R. Tambo and Tshwane representing 17.9%, 17.8% 17.8 and 17., respectively. 56.7% of the study participants were aged 15–19 while 43.3% were aged 20–24 years. Results further showed that 69.2% of the participants reported having ever had sex. Of the participants who ever had sex, 8.8% reported having started sex before or at the age of 15. Only 12.1% of the participants reported having been involved in transactional sex at some point. Table 1 shows distribution of participant's characteristics.

Table 1
Characteristics of the Study Population

Characteristic	Frequency	Weighted Percentage
District		
Cape Town	377	8.6
Ehlanzeni	803	18.3
O R Tambo	690	15.7
Tshwane	767	**17.5**
uThungulu	748	17.0
Zululand	1,014	22.1
Age Group		
15–19	2,515	57.2
20–24	1,884	42.8
Ever Had Sex		
No	1,390	31.6
Yes	3,009	68.4
Age at Sexual Debut		
≤ 15	259	5.9
16–24	2,750	62.5
Not Applicable	1,390	31.6
Highest Education		
None or Primary	210	4.8
Secondary	3,892	88.5
Some Post-Secondary	297	**6.8**
History of Transactional Sex		
Yes	424	9.6
No	4,170	90.4
Socio-economic Status		
High SES	792	18.0
Low SES	3,607	82.0

Characteristic	Frequency	Weighted Percentage
Relationship Status		
Married	39	0.9
Dating	2,693	**61.2**
Single	1,624	36.9
Divorced	3	0.1
Prefer Not to Say	40	0.9
HIV Status		
HIV Positive	568	12.9
HIV Negative	3,831	87.1
Ever Treated for STIs		
Yes	990	22.5
No	3,409	77.5
Partner HIV Status		
HIV Positive	141	3.2
HIV Negative	2,013	45.8
Prefer not to Say	86	2.0
Do not Know	899	20.4
Do not have Partner	1,260	28.6
Often Use Substances		
Yes	415	9.4
No	3,983	90.6

The overall weighted HIV prevalence among the study participants was estimated to be 12.4%. HIV prevalence was high (19.7%) among AGYW aged 20–24 compared to their 15–19 counterparts. AGYW who reported having been involved in transaction sex had higher prevalence of HIV (20.0%) compared to those who never practiced transactional sex. There was high HIV prevalence in AGYW in the low socio-economic (13.2%) compared to those in the high socio-economic status category (9.2%). The prevalence of HIV was extremely high (57.6%) for AGWY in a relationship a known HIV-positive partner. The HIV prevalence among AGYW who did not know the partner's HIV status and those who preferred not to reveal the partners status was 24%. Participants who reported having ever been treated for STIs had higher prevalence of HIV (18.5%) compared to those who were not treated for STIs (10.6%). There were no

significant differentials in HIV prevalence among AGYW who reported drug and substances misuse or not. Having six or more partners did not predict HIV in this study. Table 2 shows comparison of HIV status by selected predictor variables and their corresponding p-values.

Table 2
Comparison of HIV Status by Predictor Variables

Characteristic	HIV Positive (Weighted Row%)	HIV Negative (Weighted Row%)	P-Value
Age Group (years)			< 0.001
15–19	185 (6.75)	2,330 (93.25)	
20–24	383 (19.7)	1,501 (80.3)	
Age of Sexual Debut (years)			< 0.001
≤ 15	43 (15.9)	216 (84.1)	
16–24	456 (15.25)	2294(84.75)	
Not Applicable	69 (4.5)	1.321(95.5)	
Transactional Sex			< 0.001
Yes	90(20.0)	334 (80.0)	
No	478 (11.56)	3,497(88.44)	
Socio-economic Status			0.036
High SES	78 (9.20)	714 (90.80)	
Low SES	490 (13.20)	4,117 (86.8)	
Highest Education			0.146
None or Primary	35 (15.35)	175 (84.65)	
Secondary	501 (12.36)	3,391(87.64)	
Some Post-Secondary	32(10.31)	265(89.69)	
Partner Older than 5yrs			< 0.001
Yes	211 (20.28)	839 (78.72)	
No	304(13.92)	1,785 (86.08)	
Do not Know/NA	53(3.69)	1,207 (96.31)	
Partner HIV Status			< 0.001
HIV Positive	84(57.63)	57 (42.36)	
HIV Negative	191(8.94)	1,822(91.06)	
Prefer Not to Say	21(23.68)	65(76.32)	
Do not Know	219(24.24)	680(75.76)	

Characteristic	HIV Positive (Weighted Row%)	HIV Negative (Weighted Row%)	P-Value
Do not have Partner	53(3.69)	1,207 (96.31)	
Ever Treated for STIs			< 0.001
Yes	191 (18.45)	799(81.55)	
No	377(10.56)	3,032(89.44)	
Has more than Six Partners			0.7045
Yes	45(11.67)	299(88.33)	
No	523(12.43)	3,532 (87.57)	
One Parent Dead			< 0.001
Yes	345 (16.76)	1,688(83.24)	
No	219(9.0)	2,065(91.0)	
Often Use Substances			0.559
Yes	57(12.14)	358(87.86)	
No	511(12.39)	3,473(87.61)	

Independent Risk Scores of HIV Among AGYW

Table 3 shows coefficients from the final multivariable prediction model. Ten candidate predictors were used in the final model. Compared to AGYW aged 15–19, 20–24 years had higher risk of (B = 0.77, p < 0.001). The risk of HIV infection was slightly higher for AGYW in the lower socio-economic status (B = 0.15, p = 0.010) compared to those in high socio-economic status. Being in a relationship with a partner known to be HIV positive was associated with 11 times higher risk of being tested HIV positive (B = 2.4, p < 0.001). AGYW who did not know the status of the partner and those who preferred not to reveal the status of the partner were also associated with higher risk of being tested HIV positive (B = 1.2 and 1.0 respectively, P < 0.001). The risk score associated with AGYW who were ever treated for STIs was slightly high (0.33) compared to those who were not treated for STIs. AGYW who lost a parent had slightly higher risk score (0.80). Higher levels of education were lower risk of HIV infection compared those with primary education (B= -0.5, p = 0,007) for secondary level and (B= -1.11, p < 0.001) for post-secondary education. There were no significant differentials in the risk of HIV with respect to age at first sex. The optimal coefficient cut point estimated using youden index was 1.80 with sensitivity cut point of 62%, specificity cut point of 70% and AUC of 0.66. This shows that with a risk score of 1.80 and above, the AGYW is more likely to test HIV positive. Table 4 shows selected cut-points at various levels

Table 3
Coefficients and risk scoring for the independent predictors of HIV using the training dataset

Predictor	Coefficient	Risk Score	P-Value	95% CI	
				Lower	Upper
Age Group					
15–19	0	0			
20–24	0.7699	0.80	< 0.001	0.5487	0.9910
Age at First Sex					
≤ 15	0				
>15	0.0060	0.01	0.975	-0.3771	0.3891
Never Had	-0.1691	-0.17	0.629	-0.8551	0.5167
Relationship Status					
Single	0.2277	0.23	0.059	-0.0084	0.4639
Dating	0	0			
Married	-0.8011	-0.80	0.214	-2.0651	0.4628
Prefer Not to Say	0.1316	0.13	0.813	-0.9600	1.2233
Divorced/Separated	Omitted				
Education					
Primary	0	0			
Secondary	-0.5817	-0.58	0.007	-1.003	-0.1603
Post-Secondary	-1.1060	-1.11	< 0.001	-1.6832	– .5287
Socioeconomic Status					
Relatively Low SES	0.1528	0.15	0.297	-0.1341	0.4398
High SES	0	0			
Partner 5 years Older					
Yes	0.1734	0.17	0.124	-0.0475	0.3944
No	0	0			
No Partner	-0.3038	-0.30	0.383	-0.9864	0.3837
Ever Treated for STIs					

Predictor	Coefficient	Risk Score	P-Value	95% CI	
Yes	0.3266	0.33	0.003	0.1080	0.5452
No	0	0			
Partner HIV Positive					
Yes	2.4069	2.41	< 0.001	2.0179	2.7960
No	0	0			
Prefer not to say	1.1845	1.18	< 0.001	0.6289	1.7401
Do not Know	1.0079	1.01	< 0.001	0.7864	1.2295
No Partner	0	0			
Condom use at last sex					
Yes	0.2932	0.29	0.050	-0.0028	0.5893
No	0	0			
Transactional Sex Ever					
Yes	0.3266	0.33	0.401	-0.1655	0.4137
No	0	0			
Parent Dead					
Yes	0.8016	0.80	0.131	-0.2397	1.8430
No	0	0			

Table 4
Performance characteristics of the selected risk scores based on the training dataset

CUT POINT	SENSITIVITY	SPECIFICITY	CORRECTTLY CLASSIFIED	LR+	LR-
>=0	100.00%	0.00%	12.83%	1.0000	
>=0.5	100.00%	0.07%	12.89%	1.007	0.0000
>=1.0	94.68%	21.68%	31.05%	1.2090	0.2452
>=1.5	73.67%	39.79%	44.14%	1.2236	0.6617
>=2.0	40.765	71.20%	67.29%	1.4152	0.8320
>=2.5	21.53%	77.87%	70.64%	0.9723	1.0079
>=3.0	0.76%	99.93%	87.20%	10.1923	0.9931
>=3.5	0.00%	100.00%	87.17%	-	1.0000
The optimal cutpoint of the risk score was 1.80 with a sentivity of 62% and specificity of 70%					
Key					
Sensitivity: The proportion of actual positives which are correctly identified as such.					
Specificity: The proportion of negatives which are correctly identified as such					
LR+: The ratio of the probability of a positive test among the truly positive subjects to the probability of a positive test among the truly negative subjects.					
LR-: The ratio of the probability of a negative test among the truly positive subjects					

Performance of the Risk Prediction Model.

Table 5 shows summary of the prediction model's classification. The model's sensitivity was 15.2% with a specificity of 98.9%. The positive predictive value was 67.4% while the negative predictive value was 88.8%. The model correctly classified 88.2% of the outcome. The overall predictive power of the model at discriminating HIV positive versus HIV negative status was good as evidenced by high training and testing AUC. The training AUC was 0.770 while the testing AUC was 0.751. Figure 1 shows training and testing AUC of the prediction model. Table 6 shows summary of discrimination and calibration statistics of the prediction model. The prediction model also showed good calibration with a non-significant Hosmer Lemeshow test ($p = 0.6341$) and a brier score of 0.0933. The pseudo R^2 values of the prediction model was also high (0.1522).

Table 5
Prediction model's classification table based on the training dataset

Sensitivity	**Pr (+|D)**	**15.19%**
Specificity	Pr (-|-D)	98.92%
Positive predictive value	Pr (D|+)	67.42%
Negative predictive value	Pr (-D|-)	88.79%
False positive rate	Pr (+|-D)	1.08%
False negative rate	Pr (-|D)	84.81%
False positive rate for classified positive	Pr (-D|+)	32.58%
False negative rate for classified negative	Pr (D|-)	11.21%
Correctly classified		88.17%
X-axis is the model's sensitivity while y-axis is 1- Specificity. The blue line represents the training ROC while red line represents the testing ROC.		

Table 6
Performance of the risk prediction model on the testing dataset

Discrimination	**Value**	**Remark**
AUC	0.771	Good
Calibration	**Value**	**Remark**
Hosmer Lemeshow	0.341	Good
Brier Score	0.0933	Good
Pseudo R^2	0.1522	Good

Discussion

In this study, we aimed at developing and validating a risk prediction model for predicting elevated risk AGYW based on selected predictors of HIV infection. The overall HIV prevalence was estimated to be 12.4%. Some districts had high HIV prevalence compared to others. Our risk prediction model showed both good discrimination and calibration at predicting undiagnosed HIV. Based on the Youden Index score, an optimal risk score cut point of 1.80 may be indicative of a positive HIV status. This means that all WHO whose's risk score is above 1.80 should be screened for HIV and offered preventive services such as HIV PreP. This study has found that AGYW exposed to HIV positive partners have more than double that risk of HIV compared to HIV negative partners. This finding is not unusual because of the increased exposure to the HIV in the absence of protection and the use of PreP. Deliberate efforts are needed to

cover such AGYW on PreP to reduce the likelihood of seroconversion. Similarly, the risk score was slightly high in AGYW who were ever treated for STIs. This finding supports the idea of offering HIV testing to all clients visiting STIs to ascertain their HIV status. The optimal cut point identified in this study for both risk scores and predicted probabilities do not replace routine screening in clinical care settings. The cut points are meant to help health care workers to stratify AGYW based on risk score and provide services based on the risk score These findings inform users of the risk tool that HIV programming for AGYW should particularly target AGYW exposed to HIV positive partners, those who started sex before or at the age of 15 years and those infected with sexually transmitted infections because of their increased risk.

Based on the number of variables selected to assess risk, the risk classification can change to fit a particular user's situation. Some prediction tools are based on proximate predictors of HIV only while others include all socio-demographic and proximate predictors. According to a risk tool developed in Malawi using data from VOICE trial, only 6 variables were used to develop a risk prediction, the tool lacked key variables such as partner HIV status, age at first sex and whether the AGYW was involved in transactional sex or not (100). Similarly, an HIV risk assessments tool for AGYW in South Africa developed using data from the VOICE trial also lacked variables such as partner HIV status and transactional sex (101). Lack of key variables that are strongly predict HIV in a prediction model may reduce the tool's performance at discriminating AGYW at elevated risk for new HIV infections. This study therefore presents a more robust tool that has been developed and validated to accurately capture elevated risk AGYW. Based on behavior change and other circumstances, risk scores for AGYW can change hence need to periodically review and follow up risk over time. Studies have also shown that risk prediction models loose prediction power over time. It is therefore important to update the model by either using new data or adding candidate variables (61,95). Clinical prediction models for HIV have the potential to increase the number of AGYW to be covered on PreP to reduce their risk of acquiring HIV infection. This however depends on accuracy of the prediction model at identifying high risk AGYW. If AGYW have been falsely identified as high risk when their actual risk is low, they will be put on PreP when they are not supposed to PreP. Likewise, AGYW falsely identified as low risk when their actual risk is high will not be covered with PreP which may them get exposed to HIV infection. Minoring risk status overtime is important to prevent and correct this misclassification. Use of risk assessment tool in identifying high risk AGYW maybe a cost-effective way of offering PreP services since it may not be feasible to target all predetermined high-risk groups or based on self-assessment.

Implications and Policy Applications

Results from this study have many policy implications and applications in the area of HIV programming since evidence has shown that traditional methods have shown to be ineffective at preventing new HIV infections hence the need for a more targeted approach based on risk scoring (43). Firstly, risk scores can be used by service providers to provide health education and counseling to AGYW based on stratified risk scores, this can help save time and resources by prioritizing high risk AGYW (10, 26, 29). Secondly, HIV risk prediction tools have also been found to be useful in monitoring changes in risk over time to see if the AGYW's risk score is changing from low to high or vice versa (47).

Limitations

This study has some limitations. First the study utilized cross-sectional survey data from six South African provinces only. This can affect generalizability in other settings and countries where HIV prevalence is low. However, results from this study could potentially be relevant and be applied in countries with high HIV prevalence mostly in sub-Saharan African countries where the prevalence of HIV is high. This study targeted AGYW only, this can affect the model's applicability to adolescent boys and young men. The validation of the model was done using the same data set. The data used to develop the risk tool had 40% non-response rate, this may have the potential to affect the finding if there were systematic differences between those who accepted to take part in the survey and those who did not. This may affect applicability of the risk tool to adolescent boys and young men. There are many methods of developing risk prediction models such as generalized linear models and ensemble methods, this study used generalized linear models at the expense of other methods such as ensemble models that perform both feature selection and modeling to increase prediction accuracy. However, due good discrimination and calibration, our prediction model can perform well when applied to an external population of AGYW. The risk tool does better with more in-depth and personal exposure questions, this many affect its use in busy settings where more time is needed and when participants are not willing to disclose such information.

Recommendations

The risk tool may also be recommended in allocation of resources when there is need to allocate more resources to high-risk populations in resource limited settings. We recommend use of this risk prediction tool to identify AGYW who should be targeted with HIV PreP services to reduce risk of HIV acquisition among AGYW. A feasibility study of using the risk tool in a clinical setting to check if the tool can identify high risk AGYW to increase coverage of PreP is also recommended. Changes in prevalence of the outcome may affect the predictions models performance at predicting the outcome. There is need update the prediction with new data to reflect changes in the prevalence of outcome and its predictors. The period for updating and recalibrating the risk prediction model depends on changes on the prevalence of the outcome.

Conclusion

The developed risk score tool has shown to have high prediction accuracy at predicting AGYW at elevated risk of acquiring HIV. Such a The tool can therefore provide a simple and cost-effective way of screening AYGW women who may be targeted for receiving HIV PreP as it may be used in clinical care settings to supplement decision-making.

Abbreviations

AGYW	Adolescent Girls and Young Women

ART	Antiretroviral Therapy
AUC	Area Under Receiver Operator Characteristic Curve
CDC	Centers for Disease Control and Prevention
HIV	Human Immunodeficiency Virus
PEP	Post Exposure
PPS	Probability proportional to Size
PreP	Pre-Exposure Prophylaxis
SADHS	South African Demographic and Health Survey
SAL	Small Area Layers
SSA	Sub-Saharan Africa
STI	Sexually Transmitted Infections
UNAIDS	Joint United Nations Program on Acquired Human Immunodeficiency Syndrome
UNICEF	United Nations Children Fund
VAW	Violence Against Women
WHO	World Health Organization

Declarations

Acknowledgements

We would like to acknowledge South African Medical Research Council (SAMRC) for authorizing HERStory data to be used in the development of a risk prediction model.

Authors Contributions

Study design: Moyo R, Govindasamy D, Nyasulu PS and Manda S; Data collection and cleaning: Govindasamy D; Data analysis and interpretation: Moyo R, Manda S, Nyasulu PS and Govindasamy D; Writing of the draft manuscript: Moyo R; Read and approve the final manuscript: All authors.

Funding

This study was not funded.

Availability of Data and Materials

The data and materials used in this study may be made available upon request from SAMRC.

Ethics Approval

This study was approved by Stellenbosch University health research ethics committee (S21/09/171). There was actual involvement of humans during data collection in the initial HERStory survey. All methods and procedures during data collection were conducted in accordance with relevant guidelines and regulations of research involving humans. Informed consent was obtained from all subjects and or participants representatives.

Consent for Publication

Not Applicable

Competing Interests

The authors declare no competing interests.

Author Details

1 Reuben Moyo, Division of Epidemiology and Biostatistics, Stellenbosch University, South Africa

2 Darshini Govindasamy, Health Systems Research Unit, South African Medical Research Council, South Africa.

3 Samuel Manda, Department of Statistics, University of Pretoria, South Africa.

4 Peter Suwilanda Nyasulu, Division of Epidemiology and Biostatistics, Stellenbosch University, South Africa

References

1. United Nations Children's Fund (UNICEF). Aldolescent HIV prevention. http://data unicef/topic/hivaids/adolescents-youth-Accessed on 19/10/2020. 2020;
2. World Health Organization (WHO). Maternal, newborn child and adolescent health. . http://www. who. int/maternal_child_adolescent/topics/maternal/adolescent_HIV/en/. Accessed on 19/10/2020.

3. National Department of Health (NDoH), Statistics South Africa (Stats SA), South African Medical Research Council (SAMRC), and ICF. Pretoria, South Africa, and Rockville, Maryland, USA: NDoH, Stats SA, SAMRC and ICF. South Africa Demographic and Health Survey 2016. 2019.
4. SAMRC. Evaluation of a South African Combination HIV Prevention Programme for Adolescent Girls and Young Women. 2020;(August):128.
5. Saul J, Bachman G, Allen S, Toiv NF, Cooney C, Beamon T. The DREAMS core package of interventions: A comprehensive approach to preventing HIV among adolescent girls and young women. PLoS ONE. 2018;13(12):1–18.
6. Lakew Y, Benedict S, Haile D. Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. BMJ open. 2015;5(11):e008669.
7. Chersich MF, Rees H V. Vulnerability of women in southern Africa to infection with HIV: Biological determinants and priority health sector interventions. Aids. 2008;22(SUPPL. 4):27–40.
8. Skovdal M, Belton S. The Social Determinants of Health as they relate to children and youth growing up with HIV infection in sub-Saharan Africa. Children and Youth Services Review. 2014;45(C):1–8.
9. Underwood C, Skinner J, Osman N, Schwandt H. Structural determinants of adolescent girls' vulnerability to HIV: Views from community members in Botswana, Malawi, and Mozambique. Social Science and Medicine. 2011;73(2):343–50.
10. Buot MLG, Docena JP, Ratemo BK, Bittner MJ, Burlew JT, Nuritdinov AR, et al. Beyond race and place: Distal sociological determinants of HIV disparities. PLoS ONE. 2014;9(4).
11. Asaolu IO, Gunn JK, Center KE, Koss MP, Iwelunmor JI, Ehiri JE. Predictors of HIV testing among youth in sub-Saharan Africa: A cross-sectional study. PLoS ONE. 2016;11(10):2354500.
12. Ijumba N. Impact of HIV / AIDS on Education and Poverty About the author.
13. Fotso AS, Banjo O, Akinyemi JO. ARTICLE HIV and adolescents' educational attainment in South Africa : Disentangling the effect of infection in children and household members. 2018;
14. Impact THE, Hiv OF, Girls AON, Globally YW. IWHC THE IMPACT OF HIV AND AIDS ON GIRLS.
15. SANAC. South Africa's National Strategic Plan on HIV/Aids, TB and STI's 2017-2022. South African National AIDS Council. 2017;v.1.
16. Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA. An ambitious treatment target to help end the AIDS epidemic. Unaids. 2016;1–40.
17. Hill LM, Maseko B, Chagomerana M, Hosseinipour MC, Bekker LG, Pettifor A, et al. HIV risk, risk perception, and PrEP interest among adolescent girls and young women in Lilongwe, Malawi: operationalizing the PrEP cascade. J Int AIDS Soc. 2020;23(S3):40–7.
18. Celum CL, Delany-Moretlwe S, McConnell M, Van Rooyen H, Bekker LG, Kurth A, et al. Rethinking HIV prevention to prepare for oral PrEP implementation for young African women. Journal of the International AIDS Society. 2015;18(Suppl 3):1–10.
19. PrePWatch SA. A snapshot of PrEP scale-up, registration and resources for South Africa. 2021.

20. Wilton J, Kain T, Fowler S, Hart TA, Grennan T, Maxwell J, et al. Use of an HIV-risk screening tool to identify optimal candidates for PrEP scale-up among men who have sex with men in Toronto, Canada: Disconnect between objective and subjective HIV risk. Journal of the International AIDS Society. 2016;19(1):1–10.
21. SAMRC. Evaluation of a South African Combination HIV Prevention Programme for Adolescent Girls and Young Women. 2020;(August):128.
22. Ranganathan M, Heise L, Pettifor A, Silverwood RJ, Selin A, Macphail C, et al. Transactional sex among young women in rural South Africa: Prevalence, mediators and association with HIV infection. Journal of the International AIDS Society. 2016;19(1):1–13.
23. Price JT, Rosenberg NE, Vansia D, Phanga T, Bhushan NL, Maseko B, et al. Predictors of HIV, HIV Risk Perception, and HIV Worry Among Adolescent Girls and Young Women in Lilongwe, Malawi. Journal of acquired immune deficiency syndromes (1999). 2018;77(1):53–63.
24. Hill C, Carolina N, Carolina N, Carolina N. NIH Public Access. 2013;39(1):65–71.
25. Pettit AC, Bian A, Schember CO, Rebeiro PF, Keruly JC, Mayer KH, et al. Development and Validation of a Multivariable Prediction Model for Missed HIV Health Care Provider Visits in a Large US Clinical Cohort. Open Forum Infect Dis. 2021;8(7):1–9.
26. Rosenberg NE, Kudowa E, Price JT, Pettifor A, Bekker LG, Hosseinipour MC, et al. Identifying adolescent girls and young women at high risk for HIV acquisition: A risk assessment tool fromthe girl power-malawi study. Sex Transm Dis. 2020;47(11):760–6.
27. Pasteur L, Koch R. 1. Introduction 1. Introduction. 1941;74(1934):535–46.
28. Su TL, Jaki TF, Hickey G, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models Authors: Su, Ting-Li; Jaki, Thomas; Hickey, Graeme L; Buchan, Iain; Sperrin, Matthew. Stat Methods Med Res. 2015;1–38.
29. Jenkins DA, Martin GP, Sperrin M, Riley RD, Debray TPA, Collins GS, et al. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? Diagn Progn Res. 2021;5(1):1–7.
30. Haukoos JS, Lyons MS, Lindsell CJ, Hopkins E, Bender B, Rothman RE, et al. Derivation and validation of the Denver Human Immunodeficiency Virus (HIV) risk score for targeted HIV screening. American Journal of Epidemiology. 2012;175(8):838–46.
31. Mathur S, Pilgrim N, Patel SK, Okal J, Mwapasa V, Chipeta E, et al. HIV vulnerability among adolescent girls and young women: a multi-country latent class analysis approach. International Journal of Public Health. 2020;65(4):399–411.
32. Id HH, Yang F, Id SM, Mbita G, Bangser M, Rucinski K, et al. PLOS ONE Characterizing a sexual health and HIV risk stratification scale for sexually active adolescent girls and young women (AGYW) in Tanzania. 2021;1–15.
33. Luo Q, Huang X, Li L, Ding Y, Mi G, Scott SR, et al. External validation of a prediction tool to estimate the risk of human immunodeficiency virus infection amongst men who have sex with men. Medicine (United States). 2019;98(29):1–7.

Figures

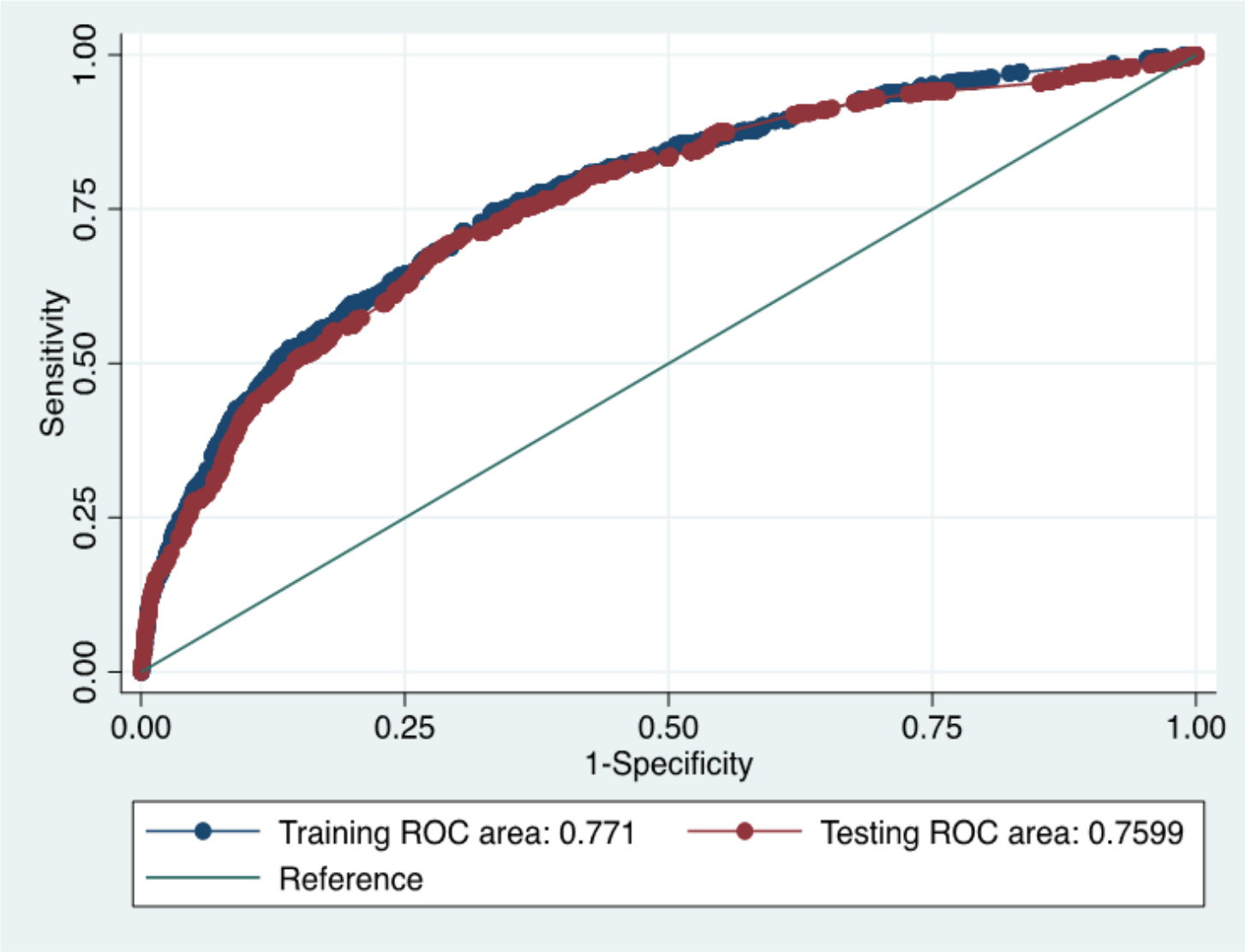


Figure 1

Training and testing AUC of the prediction model

X-axis is the model's sensitivity while y-axis is 1- Specificity. The blue line represents the training ROC while red line represents the testing ROC.