ORIGINAL RESEARCH

Development of a prognostic tool exploring female adolescent risk for HIV prevention and PrEP in rural South Africa, a generalised epidemic setting

Sarah Gabrielle Ayton [ORCID] ,[1,2] Martina Pavlicova,[2] Hod Tamir,[3] Quarraisha Abdool Karim[1,4]

[1]Department of Epidemiology, Columbia University Mailman School of Public Health, New York City, New York, USA
[2]Department of Biostatistics, Columbia University Mailman School of Public Health, New York City, New York, USA
[3]ICAP, Columbia University Mailman School of Public Health, New York City, New York, USA
[4]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa

Correspondence to
Sarah Gabrielle Ayton, Department of Biostatistics, Columbia University Mailman School of Public Health, 722 W. 168th Street, 6th floor, Rm. 635, NY 10032, USA; sarah.ayton@columbia.edu

ABSTRACT

Objectives Adolescent females in sub-Saharan Africa bear a disproportionate burden of new HIV infections but have been excluded from prognostic research, such as developed risk calculators. This study examines whether validated risk calculators, which calculate HIV risk among sub-Saharan African women, can be modified to assess HIV risk among adolescent girls. The performance of selected risk variables from validated calculators and the literature was evaluated among adolescent females using modern advanced statistical tools.

Methods Risk variables for the updated tool were selected from the CAPRISA 007 (CAP007) trial (2010–2012) questionnaires. An initially HIV-seronegative cohort of rural South African female high school students (n=1049) aged 14–25 years was selected. The number and characteristics of latent factors, or dimensions, underlying selected variables were assessed using exploratory factor analysis (EFA). The updated tool's effectiveness identifying trends in adolescent risk were assessed with latent class analysis (LCA).

Results EFA identified two key latent factors: sexual behaviour and socioeconomic risk factors. Latent sexual behaviour risk influenced contraception use (0.883), perceived HIV risk (0.691) and pregnancy (−0.384). Latent socioeconomic risk influenced low HIV knowledge (0.371), financial dependence (0.532), prior HIV testing (−0.379) and alcohol use (−0.332). Using LCA, three underlying categories of adolescent females were identified: those with no, low and high risk of HIV (1.10%, 2.26% and 2.91% 1-year seroconversion rates, respectively). Herpes simplex virus serotype-2, condom contraception, alcohol use, pregnancy and age were significantly associated with higher risk class membership, while non-condom contraception was associated with lower risk class membership.

Conclusions Adolescent females are at unequal risk of acquiring HIV. Findings suggest the updated tool captures two main facets of adolescent characteristics and may identify differential risk. This work supports further investigation to inform development of targeted differentiated interventions and efficient prognostic tools for adolescents in high-risk settings.

INTRODUCTION

Despite tremendous progress in developing HIV treatments and reducing infant acquisition, 40% of global HIV infections occur among young people aged 15–24 years (adolescents hereafter).[1] Sub-Saharan Africa is a focal point in the ongoing HIV epidemic,[2 3] in which females have an HIV acquisition rate more than double that of male peers.[4] In South Africa, approximately 30% of new infections are among adolescent females,[4 5] who are infected 5 years before males.[2 4] The disproportionate burden of HIV among females reflects and perpetuates age–sex disparate or intergenerational HIV infection, in which HIV is transmitted intergenerationally from adult male to adolescent female sexual partners.[6–8] Such relationships are prone to power imbalance and intimate partner violence, which decrease safe sex practices and increase HIV infection risk.[9 10] Notwithstanding substantial HIV burden and vulnerability,[11] adolescents have been largely excluded from groundbreaking research, discoveries and interventions.[12 13]

HIV prevention strategies for adolescents include abstinence, delay of sexual debut, consistent condom use and monogamy with HIV-seronegative partners.[6] Coupled with partner power dynamics and HIV susceptibility,[9 10] prevention strategies reliant on male partner cooperation fail to adequately prevent HIV transmission in this high priority population[14 15]; African females continue to comprise the largest proportion of new HIV infections globally.[2] Research has suggested targeted administration of pre-exposure prophylaxis (PrEP) could disrupt HIV transmission and bypass partner cooperation obstacles as an efficient and effective female-initiated prevention strategy.[16] While recent work has examined this possibility among African adult females through the development of risk calculating tools (including demographic, partner, behavioural and clinical characteristics),[17 18] South African female adolescents have been overlooked, and remain vulnerable and underserved in the ongoing South African HIV epidemic.

To effectively intervene in female adolescent population, risk calculators must be developed using adolescent-relevant HIV risk variables, identifying differential risk. While data on adolescents are scarce, the CAPRISA 007 (CAP007) trial provides a unique and valuable opportunity to explore HIV risk and behaviours among adolescent female students.[19 20] These analyses use data from one of the highest HIV burden health districts in rural KwaZulu-Natal to assess an adapted risk tool,

including characteristics from validated tools for HIV prediction in adults,[17][18] as well as adolescent risk characteristics (knowledge,[1][21] school attendance[22] and perceived risk).[23] First, we assess latent risk factors, or dimensions of adolescent risk, underlying risk variables captured with our tool. Second, we explore the differential groupings (latent subject classes) of adolescent females based on risk variables in our tool.

In this manuscript, we explore the efficacy of risk variables in predicting adolescent HIV infection to provide a foundation for enhanced prognostic medicine (risk calculators) and PrEP interventions.

METHODS
The CAPRISA 007 (CAP007) trial, a randomised controlled trial (open-label, matched-pair design), evaluated the impact of a conditional cash incentives programme on preventing HIV and herpes simplex virus serotype-2 (HSV-2) infection among 9th and 10th grade students (n=3217) in 14 rural South African high schools from 2010 to 2012. Our cohort study includes female adolescent CAP007 participants, aged 14–25 years, who enrolled in 2010, were HIV-seronegative in 2011 and followed up in 2012 (n=1049).

Study measures included HIV (HIV ELISA from Vironostika Uniform 11 plus O Assay, Biomerieux (Netherlands); HIV-positive samples were confirmed with SD Bioline HIV-1/2 ELISA V.3.0 kit (SD Standard Diagnostics, Korea)) and HSV-2 serology (HerpeSelect HSV-2 ELISA Kits (Focus Diagnostics, Cypress, California, USA) for qualitative detection of human IgG class antibodies to HSV-2 based on recombinant gG2), measured at baseline, 12 months and 24 months. Self-administered questionnaires assessed academic performance, demographic information and risk data at baseline, 12 months and 24 months. Further details about CAP007 have been published elsewhere.[19][20]

Risk variables used in analyses were derived from data collected in 2011, and HIV outcome status was obtained from 2012 data. Based on validated adult HIV risk calculators,[17][18] age category, financial support, alcohol use, contraceptive use, pregnancy, prior HIV testing and HSV-2 serostatus were included in our tool. We also incorporated risk variables unique to adolescents, including HIV knowledge,[1][21] school absence[22] and perceived HIV risk,[23] to account for potential dimensions of adolescent risk not captured by adult risk tools (online supplementary dictionary).

Statistical methods
Latent risk factors
We conducted exploratory factor analysis (EFA) to explore and gain insights into dimensions of adolescent risk captured by the updated tool. EFA explores the variance structure of correlation coefficients by analysing common variance and ignoring error variance among latent factors, represented by measured variables.[24] We ran the data through a correlation matrix, orthogonally rotated components and performed extraction with principal axis factoring; preliminary factors were extracted by variable type. The WLSMV estimator was used to analyse the tetrachoric correlation matrix. Eigenvalue assessment showed three factors with eigenvalues greater than 1.0, indicating as many as three latent factors underlying the updated tool.

We ran EFA with one, two and three factors. The χ^2 goodness-of-fit test assessed model fit. As this test is sensitive to sample size and data normality, we also examined the comparative fit index (CFI), Tucker-Lewis Index (TLI) and the root mean square error of approximation (RMSEA). Fit indices and cut

points included RMSEA (≤0.05), CFI (≥0.90) and TLI (≥0.90) to indicate good and acceptable model fit.[25][26] Indicators that appeared weak (<0.30) or displayed cross-loadings in the factorial solution were excluded from further analysis. The drug use variable was excluded from factor analyses due to low prevalence, which caused model failure to converge. The final two-factor model was chosen based on fit indices and meaningfulness of risk variable grouping. All factor analyses were performed in Mplus V.7.[27]

Latent subject classes
Latent class analysis (LCA) investigates the underlying risk classes among female adolescent participants based on measured risk variable responses.[28] This model-based clustering technique identifies subject subtypes from multivariate data and provides insight into the updated tool's ability to identify differential risk among adolescent females. The number of underlying classes, class prevalences and probability that a subject falls within a particular class may be estimated by examining associations between response and latent class. Models were fit with up to four classes, and the final number of classes was selected based on class interpretability and the sample-size adjusted Bayesian information criterion (BIC) fit statistic.[29] Since HIV status was not part of class defining variables, the frequency and proportion of HIV seroconversions was computed for each class. In the second step, we examined relationships between individual risk variables and each class using multinomial logistic regressions. This allowed us to identify risk variables significantly associated with class membership. Class analyses were performed in Mplus V.7.[27] Multinomial logistic regressions were performed in R V.3.4.1.[30]

All hypothesis tests were two-sided and assessed on the 5% level of significance.

RESULTS
Among female adolescent participants (n=1109), 20 were HIV seropositive and 111 were HSV-2 seropositive in 2011; there were 18 HIV seroconversions and 55 HSV-2 seroconversions as of 2012 (table 1). Table 1 includes descriptive statistics for these CAP007 female participants. Between 2011 and 2012, 40 HIV-seronegative female participants were lost to follow-up. A total of 1049 female adolescents met the inclusion criteria for our analysis: female participants, aged 14–25 years, enrolled in 2010, HIV-seronegative in 2011 and followed up in 2012. Of our sample, most had no past-year contraception use, high HIV knowledge, prior HIV testing, a source of spending money, no past-year drug or alcohol use, no past-year pregnancy, attributed school absences to illness or had no past-year absences, self-identified as low or no risk of HIV infection, and were under 18 years.

Latent risk factors
Based on model fit indices for the three models, geomin rotated factor loadings and parsimony, the best representation of risk variables is within the two-factor approach (online supplementary table 1). In the two-factor approach, there is a clear, meaningful loading of risk variables on two latent factors relating to HIV risk. Contraception, pregnancy and perceived HIV risk had high and significant loadings (0.883, –0.384 and 0.691, respectively) on the first factor (figure 1); HIV knowledge, prior HIV testing, financial dependence, and alcohol use had high and significant loadings (0.371, -0.379, 0.532 and -0.332, respectively) on the second factor. The first factor was termed

Table 1 Demographic information collected in 2011 among CAP007 adolescent females (enrolled in 2010) stratified by 2011 HIV serostatus

Risk variable	HIV+ (n=20) Frequency (%)	HIV− (n=1049) Frequency (%)
HIV status (2011)		
Positive	20 (100.00)	–
Negative	–	1049 (100.00)
HIV status (2012)		
Positive	20 (100.00)	18 (1.72)
Negative	–	1031 (98.28)
HSV-2 status (2011)		
Positive	4 (20.00)	90 (8.58)
Negative	12 (60.00)	890 (84.84)
HSV-2 status (2012)		
Positive	5 (25.00)	142 (13.54)
Negative	11 (55.00)	839 (79.98)
Contraception (past year)		
Condom use	7 (35.00)	270 (25.74)
Other contraception	3 (15.00)	204 (19.45)
None	10 (50.00)	575 (54.81)
HIV knowledge		
Low	–	91 (8.67)
High	20 (100.00)	958 (91.33)
Prior HIV testing		
Yes	15 (75.00)	919 (87.61)
No	5 (25.00)	121 (11.53)
Financial dependence		
Yes	3 (15.00)	114 (10.87)
No	17 (85.00)	935 (89.13)
Drug use (past year)		
At least once	1 (5.00)	38 (3.62)
Never	19 (95.00)	997 (95.04)
Alcohol use (past year)		
At least once	4 (20.00)	160 (15.25)
Never	16 (80.00)	879 (83.79)
Pregnancy		
Yes	1 (5.00)	43 (4.10)
No	19 (95.00)	1000 (95.33)
School absence (past year)		
Illness	7 (35.00)	319 (30.41)
Illness and other reason	6 (30.00)	210 (20.02)
Other reason	4 (20.00)	177 (16.87)
None	3 (15.00)	343 (32.70)
Perceived HIV risk		
High risk	2 (10.00)	78 (7.44)
Low risk	5 (25.00)	399 (38.04)
Not at risk	13 (65.00)	572 (54.53)
Age		
≥18	10 (50.00)	292 (27.84)
<18	10 (50.00)	757 (72.16)

HSV-2, herpes simplex virus serotype-2.

the sexual risk factor, as contraception use and pregnancy are directly related with sexual activity, as is perceived HIV risk. The second factor was defined as the socioeconomic risk factor, as knowledge, prior testing, financial dependence and alcohol use are all related with education, financial status and sociocultural norms. HSV-2 status and school absence had low factor loadings (<0.30) on both factors; these risk variables were dropped from analysis. We note a low correlation between sexual and socioeconomic risk factors (R=−0.133), indicating probable factor independence.

Factor loadings of the two-factor model indicated positive relationships between condom and non-condom contraception use (0.883), and high and low perceived HIV risk (0.691); conversely, past-year pregnancy (−0.384) was negatively associated with sexual behaviour risk (figure 1). Factor loadings indicated positive associations between socioeconomic risk and low HIV knowledge (0.371) and financial dependence (0.532); loadings for prior HIV testing (at least once) (−0.379) and alcohol use in the past year (at least once) (−0.332) indicated negative relationships with socioeconomic risk (figure 1). Analyses stratified by age for adolescents <18 and adolescents ≥18 yielded similar factor loadings as the unstratified factor analysis.

Latent subject classes
To assess differential risk in the population, we classified the female adolescents using LCA and found, based on the indices, the most parsimonious categorisation of female adolescents was with a three-class split of the sample (online supplementary table 2); the two-class model did not distinguish between no-risk and low-risk participants, while the four-class model further divided the high-risk group into two non-meaningful groups. Further, the three-class model had the lowest BIC, suggesting the best split.

In the three-class model, there is a clear and meaningful grouping of female adolescents, based on the distribution of risk variable proportions, into three classes relating to HIV risk: no risk, low risk and high risk (table 2). We characterise the three classes as follows: *class 1* most clearly represented those at almost no risk of HIV infection; class 1 is likely representative of younger adolescents with little to no sexual activity or drug use, and was named the 'no risk' group. Conversely, *class 2* captured those at the highest risk of HIV infection; this group includes older adolescents who have unprotected sex and a history of substance use, and was named the 'high risk' group. *Class 3* described those at low risk of HIV infection, and includes adolescents who practice safe sex and have minimal history of substance use, and was named the 'low risk' group. Interestingly, among those classified as high risk, 2.91% became HIV seropositive in the following year; 2.26% of those designated as low risk and 1.10% of those classified as no risk became HIV positive in the following year (table 2). Due to the small number of seroconversion in the sample, the differential relationship between class and seroconversion could not be statistically assessed.

In the investigation of variable associations with the three classes, multinomial logistic regression indicated HSV-2 status, alcohol use, pregnancy and age category as significantly associated with class 3, 'low risk', membership compared with class 1, 'no risk' (table 3). Risk variables significantly associated with class 3 membership were predictive of low HIV risk compared with no HIV risk. Risk variables including HSV-2 status, contraception, alcohol use, pregnancy and age category indicated significant associations with membership in class 2, 'high risk', compared with class 1 (table 3). All risk variables significantly associated with class 2 membership compared with class 1 were predictive of high HIV risk compared with no HIV risk, with the exception of use of non-condom contraception, which had a protective effect. HIV knowledge, prior HIV testing and financial dependence were not significantly associated with class categorisation.
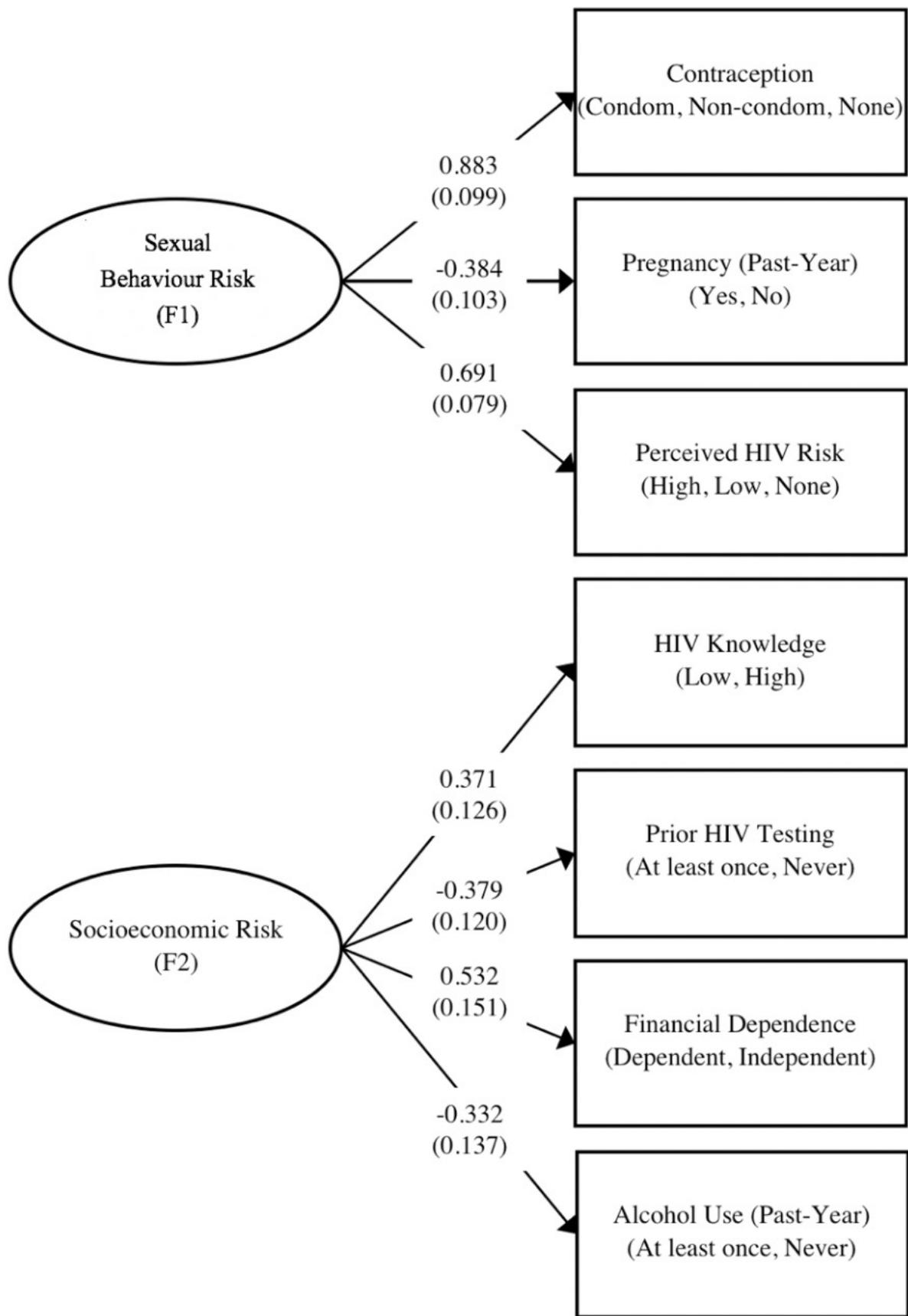
Figure 1 Geomin rotated factor loadings and standard errors of the final two-factor EFA model.

Table 2 Class risk variable proportions from the three-class LCA model among the sample of CAP007 adolescent females with HIV-negative serostatus in 2011 (n=1049)

Risk variable	Overall	Three-class model proportions		
		Class 1 'No risk'	Class 2 'High risk'	Class 3 'Low risk'
HSV-2 status				
Positive	0.092	0.052	0.234	0.096
Negative	0.908	0.948	0.766	0.904
Contraception				
Condom use	0.257	0.064	0.304	0.489
Other contraception	0.194	0.000	0.000	0.511
None	0.548	0.936	0.696	0.000
HIV knowledge				
Low	0.087	0.080	0.073	0.100
High	0.913	0.920	0.927	0.900
Prior HIV testing				
Yes	0.884	0.883	0.900	0.879
No	0.116	0.117	0.100	0.121
Financial dependence				
Yes	0.109	0.110	0.086	0.115
No	0.891	0.890	0.914	0.885
Drug use (past year)				
At least once	0.037	0.000	0.137	0.048
Never	0.963	1.000	0.863	0.952
Alcohol use (past year)				
At least once	0.154	0.068	0.457	0.158
Never	0.846	0.932	0.543	0.842
Pregnancy (past year)				
Yes	0.041	0.014	0.037	0.078
No	0.959	0.986	0.963	0.922
School absence (past year)				
Illness	0.304	0.325	0.310	0.276
Illness and other reason	0.200	0.183	0.177	0.231
Other reason	0.169	0.124	0.328	0.170
None	0.327	0.369	0.185	0.323
Perceived HIV risk				
High risk	0.074	0.085	0.249	0.000
Low risk	0.380	0.000	0.000	1.000
Not at risk	0.545	0.915	0.751	0.000
Age				
≥18	0.278	0.163	0.451	0.366
<18	0.722	0.837	0.549	0.634
n (%)	1049 (100)	547 (52.15)	103 (9.82)	399 (38.04)
HIV status (2012)				
Positive		6 (1.10)	3 (2.91)	9 (2.26)
Negative		541 (98.90)	100 (97.09)	390 (97.74)

HSV-2, herpes simplex virus serotype-2; LCA, latent class analysis.

DISCUSSION

In South Africa, safe sex educational interventions have not been dramatically successful mostly due to reliance on partner cooperation and sexual relationship equality[9 10 14 15]; few studies have examined appropriate interventions for adolescent females. Recent work has indicated that PrEP,[12 16 17] the only female-initiated HIV prevention technology, is a viable means of reducing the HIV/AIDS epidemic in regions, such as South Africa. By identifying adolescent females that are high risk for HIV acquisition, preventative treatment and interventions (ie, PrEP) may be initiated through ongoing programmes, clinics and primary care providers to mediate the burden of HIV acquisition among sub-Saharan African females.

This is the first study, to our knowledge, assessing risk variables in South African adolescent females, using exploratory factors and latent classes. We specifically assess risk variables relevant to the adolescent population through an updated specialised tool. CAP007 provided the unique opportunity to explore risk variables, adapted from validated adult HIV risk calculators, in the tool and their association with HIV acquisition risk at 1 year. In a region with high epidemic burden, we find that adolescent females are at unequal risk of HIV acquisition and experience vulnerabilities distinct from their adult counterparts.

Table 3 Univariable multinomial logistic analysis of latent class membership among CAP007 adolescent females with HIV-negative serostatus in 2011 (n=1049)

Risk variable	'Low risk' vs 'no risk' (Ref) (n=399)			'High risk' vs 'no risk' (Ref) (n=103)		
	β (SE)	P value	OR (95% CI)	β (SE)	P value	OR (95% CI)
HSV-2 status		<0.0001			<0.0001	
Positive	0.70 (0.27)	0.0087	2.02 (1.20 to 3.42)	2.16 (0.30)	<0.0001	8.64 (4.79 to 15.59)
Negative	–	–	Ref	–	–	Ref
Contraception (past year)		<0.0001			<0.0001	
Condom use	13.97 (19.62)	0.4764	1.17e+06 (2.34e−11 to 5.85e+22)	2.37 (0.27)	<0.0001	10.72 (6.33 to 18.18)
Other contraception	24.57 (39.25)	0.5314	4.67e+10 (1.82e−23 to 1.20e+44)	−4.93 (0.03)	<0.0001	0.01 (0.01 to 0.01)
None	–	–	Ref	–	–	Ref
HIV knowledge		0.4378			0.4378	
Low	−0.24 (0.23)	0.291	0.79 (0.50 to 1.23)	0.18 (0.42)	0.6661	1.20 (0.52 to 2.74)
High	–	–	Ref	–	–	Ref
Prior HIV testing		0.7658			0.7658	
Yes	0.00 (0.20)	0.9997	1.00 (0.67 to 1.49)	−0.49 (0.39)	0.2126	0.61 (0.29 to 1.32)
No	–	–	Ref	–	–	Ref
Financial dependence		0.5206			0.5206	
Yes	0.06 (0.21)	0.7875	1.06 (0.70 to 1.59)	−0.38 (0.39)	0.3326	0.68 (0.32 to 1.48)
No	–	–	Ref	–	–	Ref
Drug use (past year)		<0.0001			<0.0001	
At least once	7.12 (6.71)	0.2887	1231.57 (0.00 to 6.30e+08)	8.62 (6.71)	0.1987	5553.30 (0.01 to 2.85e+09)
Never	–	–	Ref	–	–	Ref
Alcohol use (past year)		<0.0001			<0.0001	
At least once	1.09 (0.23)	<0.0001	2.98 (1.91 to 4.66)	3.33 (0.28)	<0.0001	27.83 (16.24 to 47.72)
Never	–	–	Ref	–	–	Ref
Pregnancy (past year)		<0.0001			<0.0001	
Yes	1.86 (0.42)	0.0022	6.43 (2.80 to 14.77)	1.36 (0.60)	<0.0001	3.90 (1.21 to 12.53)
No	–	–	Ref	–	–	Ref
School absence (past year)		<0.0001			<0.0001	
Illness	−0.04 (0.17)	0.7993	0.96 (0.69 to 1.33)	0.87 (0.33)	0.0079	2.40 (1.26 to 4.58)
Illness and other reason	0.35 (0.18)	0.0563	1.42 (0.99 to 2.03)	0.87 (0.37)	0.0188	2.39 (1.16 to 4.94)
Other reason	0.39 (0.20)	0.0556	1.48 (0.99 to 2.20)	1.96 (0.33)	<0.0001	7.10 (3.68 to 13.69)
None	–	–	Ref	–	–	Ref
Perceived HIV risk		<0.0001			<0.0001	
High risk	−3.58 (434.29)	0.9934	0.03 (0.00 to Inf)	1.38 (0.27)	<0.0001	3.98 (2.37 to 6.70)
Low risk	24.73 (32.45)	0.4460	5.51e+10 (1.30e−17 to 2.32e+38)	−5.01 (738.00)	0.9946	0.01 (0.00 to Inf)
Not at risk	–	–	Ref	–	–	Ref
Age		<0.0001			<0.0001	
≥18	1.13 (0.16)	<0.0001	3.09 (2.28 to 4.21)	2.01 (0.23)	<0.0001	7.49 (4.75 to 11.79)
<18	–	–	Ref	–	–	Ref

HSV-2, herpes simplex virus serotype-2; Ref, referent.

When exploring risk variables appropriate to female adolescents, two factors emerged as underlying risk variables in the updated tool: sexual behaviour risk factor (contraception, pregnancy and perceived HIV risk) and socioeconomic risk factor (HIV knowledge, prior HIV testing, financial dependence and alcohol use). These results are intuitive, as contraception and

pregnancy are directly linked with sexual activity, and perceived risk is likely associated with internalised risk regarding sexual activity. Further, knowledge, prior testing and financial dependence are known socioeconomic risk factors for HIV transmission, and substance use patterns are directly related with HIV transmission. The role of adolescent characteristics in HIV risk factors is distinct from the adult population[6 17 18 21] and suggests these risk variables encompass multiple dimensions of latent HIV risk among adolescents.

This study identified associations between adolescent characteristics and HIV risk distinct from previous studies of risk in adults.[17 18 21] The negative relationship between past-year pregnancy and latent sexual behaviour risk factor in adolescents is opposite to what we expect among adults, as pregnancy influences biological susceptibility to HIV infection.[17] Pregnant adolescents may live with family rather than their sexual partner(s) and experience reduced sexual activity (opportunity for sexual transmission) during pregnancy. Prior HIV testing and past-year alcohol use had negative relationships with latent socioeconomic risk. Among adults, prior testing approximates behavioural responses to internalised risk and is associated with higher HIV risk.[18] This was not observed among adolescents and may represent positive proactive sexual health behaviours rather than reactive behaviours. Prior testing may result from sexual autonomy and higher socioeconomic status among adolescents. Alcohol use is associated with increased risk of HIV in adults.[21] However, due to low rates of drug and alcohol use, alcohol use may approximate protective effects of age on socioeconomic risk among adolescents.

Following identification of latent risk factors, adolescents were classified into three latent classes: class of those with essentially no risk, class of those with low risk and class of those with high risk of acquiring HIV. HIV seroconversion prevalences indicated preliminary support of classes identified by LCA; while rate of seroconversion increased steadily with risk class membership, statistical tests could not be performed. Multinomial logistic regression analyses indicated non-condom contraception protected against HIV risk, while condom use predicted increased risk. This contradicts the association observed in adults[6] and may indicate non-condom contraception among adolescents is an indicator of sexual autonomy, high socioeconomic status or partner monogamy, which protect against HIV. Conversely, use of non-condom contraception may result from fear or uncooperative partners; non-condom contraception may instead be driven by some other characteristic driving reduced HIV risk.

Limitations
We were unable to statistically assess class differences in seroconversion rates. Findings from such analyses would have been clouded by type I error and cluster effects as an artefact of the small number of seroconversions observed in the data. Future research should assess the updated tool in larger datasets to further determine its ability to distinguish risk among adolescents.

It is noteworthy that the 'no risk' group from latent class analysis showed seroconversions, indicating some degree of HIV risk is not captured in the data; there may be unmeasured variables or changes in measured variables between survey administration and HIV testing at 1 year. Future work should examine changes in variables over time along with changes in HIV status. Until such progress in identifying HIV risk has been made, it is important that those considered 'low risk' still be included in HIV prevention interventions.

Further, it is possible that pregnant and HIV-infected students are more likely to drop out of school; dynamics influencing school enrolment may have contributed to the low prevalence of HIV in the dataset, which may not fully represent the adolescent population. However, since most South African adolescents are enrolled in school for at least the first 2 years of high school, our analysis provides an opportunity to understand HIV risk trajectory among high school students.

CONCLUSIONS
Our research assessing the updated risk tool recognises clear differences in HIV vulnerability among adolescent females, as a population distinct from adults. Our examination of adolescent HIV risk, latent factor structure and latent classes of South African female adolescents provide insights into HIV susceptibility and the performance of our updated tool among female adolescents. This work is a first step towards the development of specialised risk tools for female adolescents and recognises their unique role in HIV transmission. As the HIV epidemic continues in South Africa, we underscore the importance of disrupting transmission before the epidemic can reach younger generations, while also protecting the most vulnerable. Such epidemic disruption may use advances in risk calculators, such as ours, to identify adolescents at high risk of HIV infection and provide PrEP and preventive interventions. We hope this and future research will prioritise the adolescent population, and aid in the development of prognostic tools that predict HIV risk and administer PrEP among adolescent females to end the HIV epidemic in South Africa as a public health threat.

Patient and Public Involvement
It was not appropriate or possible to involve patients or the public in this work.

> **Key messages**
>
> ► South African female adolescents are particularly vulnerable to HIV infection and display distinct risk factors from adult counterparts.
> ► The updated risk tool captures latent sexual and socioeconomic risk factors, and may identify differential risk among adolescent females.
> ► Comprehensive prevention research on adolescent females and their HIV vulnerabilities is needed to further develop specialised risk tools.

the study, revised the manuscript, and critically revised and approved the final manuscript.

ORCID iD
Sarah Gabrielle Ayton http://orcid.org/0000-0001-5247-9912

REFERENCES

1 The Joint United Nations Programme on HIV and AIDS (UNAIDS). *The GAP report*. Geneva: Joint United Nations Programme on HIV/AIDS, 2014: 123.
2 The Joint United Nations Programme on HIV and AIDS (UNAIDS). *UNAIDS report on the global HIV epidemic*, 2013.
3 Abdool Karim SS, Churchyard GJ, Karim QA, *et al*. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet* 2009;374:921–33.
4 Shisana O, Rehle T, Simbayi LC, *et al*. *South African national HIV prevalence, incidence and behaviour survey, 2012*. Cape Town: HSRC Press, 2014.
5 Cowan F, Pettifor A. Hiv in adolescents in sub-Saharan Africa. *Curr Opin HIV AIDS* 2009;4:288–93.
6 Dellar RC, Dlamini S, Karim QA. Adolescent girls and young women: key populations for HIV epidemic control. *J Int AIDS Soc* 2015;18(2 Suppl 1):19408.
7 Abdool Karim Q, Abdool Karim SS. *HIV/AIDS in South Africa*. Cambridge: Cambridge University Press, 2010.
8 Pettifor A, Macphail C, Rees H, *et al*. HIV and sexual behavior among young people: the South African paradox. *Sex Transm Dis* 2008;35:843–4.
9 Jewkes RK, Dunkle K, Nduna M, *et al*. Intimate partner violence, relationship power inequity, and incidence of HIV infection in young women in South Africa: a cohort study. *Lancet* 2010;376:41–8.
10 Seedat M, Van Niekerk A, Jewkes R, *et al*. Violence and injuries in South Africa: prioritising an agenda for prevention. *Lancet* 2009;374:1011–22.
11 World Health Organization (WHO). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations, 2016. Available: http://www.who.org;
12 Bekker L-G, Gill K, Wallace M. Pre-exposure prophylaxis for South African adolescents: what evidence? *S Afr Med J* 2015;105:907.
13 Harrison A, Newell M-L, Imrie J, *et al*. HIV prevention for South African youth: which interventions work? A systematic review of current evidence. *BMC Public Health* 2010;10:102.
14 Mavedzenge SN, Luecke E, Ross DA. *Effectiveness of HIV prevention, treatment and care interventions among adolescents: a systematic review of systematic reviews. UNICEF technical brief*. New York: UNICEF, 2013.
15 Napierala Mavedzenge SM, Doyle AM, Ross DA, *et al*. HIV prevention in young people in sub-Saharan Africa: a systematic review. *J Adolesc Health* 2011;49:568–86.
16 Govender EM, Mansoor LE, Abdool Karim Q. Influences of geo-spatial location on pre-exposure prophylaxis use in South Africa: positioning microbicides for better product uptake. *AIDS Care* 2017;29:734–40.
17 Pintye J, Drake AL, Kinuthia J, *et al*. A risk assessment tool for identifying pregnant and postpartum women who may benefit from preexposure prophylaxis. *Clin Infect Dis* 2017;64:751–8.
18 Balkus JE, Brown E, Palanee T, *et al*. An empiric HIV risk scoring tool to predict HIV-1 acquisition in African women. *J Acquir Immune Defic Syndr* 2016;72:333–43.
19 Humphries H, Kharsany ABM, Leask K, *et al*. *The impact of conditional cash transfers in reducing HIV in adolescent girls and boys (RHIVA): the CAPRISA 007 matched pair, cluster randomised controlled trial*. The CAPRISA Clinical Trials: HIV Treatment and Prevention, 2017: 77–89.
20 Abdool Karim Q, Kharsany ABM, Leask K, *et al*. Prevalence of HIV, HSV-2 and pregnancy among high school students in rural KwaZulu-Natal, South Africa: a bio-behavioural cross-sectional survey. *Sex Transm Infect* 2014;90:620–6.
21 Mathur S, Wei Y, Zhong X, *et al*. Partner characteristics associated with HIV acquisition among youth in Rakai, Uganda. *J Acquir Immune Defic Syndr* 2015;69:75–84.
22 Hallfors D, Cho H, Rusakaniko S, *et al*. Supporting adolescent orphan girls to stay in school as HIV risk prevention: evidence from a randomized controlled trial in Zimbabwe. *Am J Public Health* 2011;101:1082–8.
23 Naidoo P, Donenberg G, Davids A, *et al*. Exploring risk and protective mechanisms associated with HIV infection among adolescents in South Africa. *J Psychol Afr* 2014;24:232–40.
24 Hurley AE, Scandura TA, Schriesheim CA, *et al*. Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. *J Organ Behav* 1997;18:667–83.
25 Yu C. *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles: University of California, 2002.
26 Hu Li-tze, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1–55.
27 Muthén LK, Muthén BO. *Mplus user's guide. Eighth edition*. Los Angeles, CA: Muthén & Muthén, 1998-2017.
28 Porcu M, Giambona F. Introduction to latent class analysis with applications. *J Early Adolesc* 2017;37:129–58.
29 Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol Methods* 2007;12:121–38.
30 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.