AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 30, Number 10, 2014

DOI: 10.1089/aid.2014.0110

HIV Infection in High School Students in Rural South Africa: Role of Transmissions Among Students

Ayesha B.M. Kharsany,[1] Thulasizwe John Buthelezi,[1] Janet A. Frohlich,[1] Nonhlanhla Yende-Zuma,[1] Natasha Samsunder,[1] Gethwana Mahlase,[2] Carolyn Williamson,[1,3,4] Simon A. Travers,[5] Jinny C. Marais,[3] Rachael Dellar,[1] Salim S. Abdool Karim[1,6] and Quarraisha Abdool Karim[1,6]

Abstract

In South Africa, adolescents constitute a key population at high risk of HIV acquisition. However, little is known about HIV transmission among students within schools. This study was undertaken to assess the risk factors for HIV infection and the extent of transmission among rural high school students. Between February and May 2012, consenting students from five randomly selected public sector high schools in rural KwaZulu-Natal participated in an anonymous cross-sectional survey. Dried blood spot samples were collected and tested for HIV. β-Human chorionic gonadotropin (βHCG) levels were measured in females for pregnancy. Family circumstances as well as sociodemographic and behavioral factors were assessed as potential risk factors. A subset (106/148, 72%) of HIV-positive samples underwent gag p17p24 sequencing for phylogenetic analysis. A total of 3,242 students (81.7% of enrolled students) participated. HIV prevalence was 6.8% [95% confidence interval (CI) 3.9–9.8%] in girls and 2.7% (CI 1.6–3.8%) in boys [adjusted odds ratio (aOR) = 3.0, CI 2.4–3.8; $p < 0.001$]. HIV prevalence increased from 4.6% (95% CI 1.9–7.3) in the 12- to 15-year-old girls to 23.1% (95% CI 7.7–38.5) in girls over 20 years, while in boys HIV prevalence increased from 2.7% (95% CI 0.6–4.9) in the 12- to15-year-old boys to 11.1% (95% CI 2.7–19.4) in those over 20 years. Sequencing of samples obtained from students revealed only two clusters, suggesting within-school transmission and three interschool clusters, while the remainder was most likely acquired from sources other than those currently found in students attending the school concerned. HIV prevalence in both girls (aOR = 3.6, CI 2.9–4.5; $p < 0.001$) and boys (aOR = 2.8, CI 1.2–6.2; $p = 0.01$) was higher in those without a living biological mother. The high burden of HIV infection among students was not associated with intraschool transmission in this rural setting. Lack of a living parent is an important factor defining high risk in this group of adolescents.

Introduction

Annual, anonymous HIV testing of pregnant women[1] has been central to understanding the evolving HIV epidemic in generalized epidemic settings. As epidemics mature in these settings and with increasing coverage of antiretroviral treatment (ART) and interventions to prevent vertical transmission of HIV the prevalence of HIV infection is a less reliable marker of the evolving epidemic. Thus, there is a need to expand current surveillance systems to include new adolescent populations that provide a better marker of new HIV infections.

In South Africa, heterosexual transmission accounts for more than 90% of new infections. HIV prevalence in adolescents between 15 and 24 years in high disease burden communities provides a reasonable proxy for incident HIV infections, because infections are likely to be relatively recent and HIV-related mortality is likely to be minimal.[2–4]

[1]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Congella, Durban, South Africa.
[2]Zimnadi Zonke, Pietermaritzburg, South Africa.
[3]Division of Medical Virology, Clinical Laboratory Sciences and Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa.
[4]National Health Laboratory Service, Durban, South Africa.
[5]South African National Bioinformatics Institute, South African Medical Research Council Bioinformatics Unit, University of the Western Cape, Bellville, South Africa.
[6]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York.

The burden of HIV in the South African adolescent population is unprecedentedly high, particularly in adolescent girls aged 15–19 years who acquire HIV at least 5–7 years earlier than their male peers and have a 3- to 4-fold higher incidence rate.[5] Given that the majority of adolescents in these settings attend high school, high schools provide a convenient opportunity for expanded surveillance. Furthermore, schools also provide important venues for intensifying HIV prevention efforts and surveillance in this setting, and enable the impact of interventions to be assessed.

Phylogenetic methods have also been applied to investigate HIV transmission dynamics and are useful for identifying high-risk physical environments and providing insights into understanding HIV spread.[6] With modern analytical approaches, these important tools not only enable transmission clusters to be inferred, but also allow estimation of the timing of HIV transmission, thereby permitting smarter targeting of HIV prevention efforts.

We assessed the potential for including at-risk adolescent populations through school-based surveillance, together with the utility of phylogenetic mapping, as measures to enhance the understanding of the evolving epidemic and transmission dynamics in a high disease burden rural district of South Africa.

Materials and Methods

Study setting and population

The study was conducted in rural Vulindlela, a subdistrict of Umgungundlovu 150 km west of Durban in the province of KwaZulu-Natal, which is one of four highest burden HIV districts in South Africa. The HIV prevalence among pregnant women attending public sector primary health care clinics in Vulindlela in 2011 was 39.8%.[1] The area is characterized by high levels of poverty due to its limited resources and employment opportunities. The estimated population is about 150,000, of which approximately 16,000 are in high school.[7] The establishment of the school-based surveillance was preceded by extensive consultative meetings with key stakeholders including the Departments of Education and Health at the provincial, district, and school level and with school principals and educators, governing bodies, parents, and students.

Study design and procedures

Between February and May 2012, we conducted a cross-sectional study in five randomly selected public sector high schools within a 15 km radius of each other. Consenting students aged 12 years and older were eligible for inclusion in this survey. Prior to study initiation, parents and/or guardians of students were notified and had an opportunity to determine whether to allow or refuse their child's participation in the study. Students were provided with general information on HIV and on concepts of HIV surveillance, study information, study procedures, and the confidentiality of their participation and the data collected. Thereafter students agreeing to participate were individually provided with details of the study and taken through the informed consent process. A brief reason was obtained from students not agreeing to participate. Students consenting to study participation had dried blood spot (DBS) samples collected by trained staff who also administered a standardized structured questionnaire to obtain sociodemographic and select behavioral information. Students were assured of their anonymity both in data collection and analysis. Laboratory data were linked to sociodemographic and behavioral data through a unique code. Students had access to HIV counseling and testing (HCT), sexual reproductive health services (SRH), and medical male circumcision (MMC) services that were provided through the primary health care clinics and the CAPRISA Clinical Research site in the district; these services were also available to other family members.

HIV testing was performed using the Vironostika HIV Uni-Form II plus O Assay (Biomérieux, Netherlands) and the Elecsys HIV Combi assay (Roche Diagnostics GmbH, Germany). β-Human chorionic gonadotropin (βHCG) was measured using the Quickvue one-step HCG combo (Quidel Corporation, USA), with all positives confirmed with HCG combo rapid test (CTK Biotech, Inc., USA).

HIV-positive DBS samples (106 samples from this study and 14 samples collected previously[7]) were used for phylogenetic analysis. Using the Harris UNI-CORE 3.00-mm punch, DBS spots were punched out and nucleic acid was extracted from the samples using the QIAamp DNA Mini kit (Qiagen, Germany) according to the manufacturer's instructions, with the following modifications: a maximum of seven, 3-mm-diameter punches was processed per tube; for elution of nucleic acids, a minimum volume of 50 μl molecular biology grade water was added to the column and incubated in the column for 5 min before centrifugation. Amplicons ($\pm$750 bp) for sequencing were generated using one-step reverse transcription polymerase chain reaction (RT-PCR) and nested PCR (a section of gag p17p24 was amplified),[8,9] with the following differences[8]: 0.8$\times$ reaction mix was used (including 0.16 mM each dNTP and 0.96 mM $MgSO_4$) and 4–27 μl nucleic acids was used per RT-PCR reaction. Five microliters of prenested product was inoculated into nested PCR mixes. Prenested PCR primers used were DT1 5′ ATG GGT GCG AGA GCG TCA GTA TT 3′ (790–812 HXB2) and DT7 5′ CCC TGA CAT GCT GTC ATC ATT TCT TCT 3′ (1818–1844 HXB2) and nested PCR primers were DT3 5′ CAT CTA GTA TGG GCA AGC AGG GA 3′ (886–908 HXB2) and DT6 5′ ATG CTG ACA GGG CTA TAC ATT CTT AC 3′ (1609–1634 HXB2).[9] PCR products were sequenced using a BigDye Terminator V3.1 sequencing kit (Applied Biosystems, USA) and sequencing primers used were DT3 and DT6. Chromatograms were assembled using Sequencher software (Gene Codes, USA). For superimposed peaks in sequence chromatograms, the dominant base was used; if it was not clear which was dominant, the ambiguous code was used. Sequences were approximately 605–635 bp in length.

Ethical considerations

The protocol was approved by the Biomedical Research Ethics Committees of the University of KwaZulu-Natal (Reference number E179/04) and the KwaZulu-Natal Departments of Health and Education. The HIV sequencing protocol was approved by the University of Cape Town (Reference number 242/2011).

Data management and statistical analysis

Data were collected on standardized case report forms (CRFs), faxed to a dedicated study database using DataFax

(Clinical DataFax Systems Inc., Hamilton, Canada), and linked to laboratory data. The select demographic, behavioral, and biological data were summarized using descriptive summary measures, expressed as means [± standard deviation (± SD)] and/or medians [interquartile range (IQR)] for quantitative variables and percentages for categorical variables. The 95% confidence intervals (CI) and *p*-values ($p < 0.05$ were considered as significant) are reported for the school adjusted HIV prevalence. Comparisons of adjusted summaries by gender were performed using *t*-tests for two independent samples. The association between HIV and demographic, behavioral, and biological (pregnancy) measures was evaluated with the multivariable logistic regression analysis adjusted for clustering to estimate the odds ratios [adjusted odds ratio (aOR)] and 95% CI. All analyses were performed using SAS statistical package (version 9.3; Statistical Analysis Software, NC, USA).

Sequences were subtyped using the jumping profile hidden Markov model (jpHMM) (http://jphmm.gobics.de/)[10] and REGA HIV Subtyping Tool (http://bioafrica.net/rega-genotype/html/subtypinghiv.html).[11,12] Multiple sequence alignments were generated using RAMICS, a codon-based multiple sequence alignment tool,[13] and manually checked for accuracy. Codon-sized columns containing a majority of gaps (present in greater than 90% of sequences) were deleted from the alignment. HIV-1 subtype C gag p17p24 sequences ($n = 135$) derived from recently infected women and students ($n = 14$)[7,14–16] from previous studies in the Durban and Vulindlela areas were used. Phylogenetic reconstruction was performed using two maximum likelihood approaches implemented in FastTree (GTR + CAT + GAMMA model)[17] and RAxML (GTRGAMMA model with 100 bootstrap replicates). Putative transmission clusters were characterized as groupings with support values greater than 90% in both of the resulting phylogenies.[18]

Results

A total of 3,242/3,967 students (81.7%) (range: 75–88%) participated in the study; 365 (9.2%) refused to participate and 361 (9.1%) were absent from school at the time of the survey.

Demographic and behavioral characteristics

The median age of girls (15 years; IQR 14–17) and boys (16 years; IQR 14–17) was similar ($p = 0.60$). The demographic and behavioral characteristics are presented overall and by gender in Table 1. The proportion of girls and boys across grades 8 to 12 was similar; attrition rates were high in both girls and boys in grades 11 and 12. Family structures tended to be unstable, with almost a quarter (22.9%) of students not having a mother alive, and a further 22.6% having a biological mother alive but not living with them. Fathers were generally absent in the students' lives; indeed, only 58.3% of students reported that their biological father was alive and only 51.7% lived with their fathers.

Almost a third of the boys (33.0%) and about a fifth of the girls (18.8%) reported ever having had sex. The median age at sexual debut was 17 years (IQR 16–18) for girls and 16 years (IQR 15–17) for boys. Boys reported a higher median total number of lifetime sex partners (median 2, IQR 1–4) compared to girls (median 1, IQR 1–2). Among those who reported ever having had sex, boys (74%) compared to girls (32.4%) were more likely to have had a sex partner either the same age or younger than themselves ($p < 0.001$). However, girls were more likely to have had a partner at least 1–4 years (55.1% versus 22.6%; $p < 0.0001$) or 5 years or older (12.5% versus 2.9%; $p = 0.003$) than themselves. Pregnancy prevalence was 3.3%.

HIV infection

HIV prevalence and demographic, behavioral, and biological correlates of HIV risk are reported in Table 2. Overall, HIV prevalence was 6.8% (95% CI 3.9–9.8) in girls and 2.7% (95% CI 1.6–3.8) in boys ($p = 0.007$). There was substantial variability in HIV prevalence between schools, by sex and age, with an intraclass correlation (ICC) of 0.005 and a design effect of 4.2. In boys, HIV prevalence ranged from 1.8% in school A to 4.1% in school C. In girls, HIV prevalence was consistently higher in all schools, ranging from 3.5% in school E to 9.0% in school C.

In girls, HIV prevalence increased in a dose relationship with age from 4.6% (95% CI 1.9–7.3) in 12 to 15 year olds, to 5.3% (95% CI 0.8-9.9) in 16 to 17 year olds, 8.7% (95% CI 0–17.7) in 18 to 19 year olds, and 23.1% (95% CI 7.7–38.5) in 20 year olds or older. In contrast, in boys HIV prevalence was relatively constant at around 2.7% (95% CI 0.6–4.9) until age 19 years, but increased dramatically to 11.1% (95% CI 2.7–19.4) in those 20 years and older.

Phylogenetic relationship of HIV-1 sequences

Of the 120 samples sequenced (this study $n = 106$, $n = 14$[7]), 119 were classified as subtype C and one as A1 (subtype A1 was excluded from analysis). Figure 1 shows the reconstructed phylogenetic tree of HIV 1-subtype C groupings obtained from students in comparison to previously characterized sequences. Nine sequence clusters exclusively containing sequences from students were identified, with two samples per cluster. In four clusters (Fig. 1; clusters 1, 4, 6, and 8), two sequences in each cluster were from the same individual but were collected at different time points. The remaining five clusters (Fig. 1; clusters 2, 3, 5, 7, and 9) contained sequences from different students. Three of the five clusters were sequences from students of the same gender (two female only clusters) (Fig. 1; clusters 3 and 9), and one male only cluster (Fig. 1; cluster 7) indicating a third unidentified source of infection. In comparison to the students, clustering was observed for only 7 of the 135 background data sequences (four clusters, Fig. 1; green and gray clusters), one containing a sequence from a student (Fig. 1; green cluster), suggesting that a larger number of students in schools clustered together when computed against the background community data ($p = 0.02$). Of the five clusters, two sequences were from individuals from the same school (Fig. 1; clusters 3 and 9) and three from different schools (Fig. 1; clusters 2, 5, and 7).

Factors associated with HIV infection

The association of demographic, behavioral, and biological factors with HIV is reported in Table 3. Compared to boys, girls were three times more likely to be HIV positive, aOR = 3.0, 95% CI 2.4–3.8; $p < 0.0001$, and those reporting

Table 1. Select Demographic and Behavioral Characteristics of High School Students in Rural Kwazulu-Natal, South Africa

	Total (N = *3242*)		*Girls* (N = *1698*)		*Boys* (N = *1543*)		
Characteristic	*Unadjusted*[a]	*Adjusted*[a,b]	*Unadjusted*[a]	*Adjusted*[a,b]	*Unadjusted*[a]	*Adjusted*[a,b]	p-*value*[c]
Demographic							
Age (mean; ± SD; range)	15.8 ± 2.3; 12–28		15.7 ± 2.3; 12–28		16 ± 2.2; 12–26		
Age (median, IQR)	16 (14–17)		15 (14–17)		16 (14–17)		0.60
Age groups, % (*n*)							
12–15 years	48.4 (1,568)	46.6	52.0 (882)	50.1	44.5 (686)	43.2	0.18
16–17 years	30.6 (991)	31.3	29.0 (492)	30.2	32.4 (499)	32.6	0.29
18–19 years	14.1 (456)	14.7	12.1 (206)	12.5	16.2 (250)	16.9	0.12
≥ 20 years	6.9 (223)	7.3	6.9 (117)	7.2	6.9 (106)	7.4	0.95
Grade distribution, % (*n*)							
Grade 8	22.6 (732)	22.2	21.1 (359)	20.9	24.2 (373)	23.9	0.42
Grade 9	22.5 (727)	22.2	21.8 (369)	21.5	23.2 (358)	23.0	0.62
Grade 10	24.4 (791)	23.2	24.4 (413)	22.8	24.5 (378)	23.5	0.82
Grade 11	16.8 (603)	19.5	19.7 (334)	20.4	17.5 (269)	18.5	0.45
Grade 12	11.8 (383)	12.9	13.0 (220)	14.5	10.6 (163)	11.1	0.31
Family characteristics, % (*n*)							
Biological mother alive	77.1 (2,494)	75.8	77.8 (1,318)	76.8	76.3 (1,176)	75.1	0.57
Always lives with biological mother[d]	77.4 (1,914)	77.2	78.1 (1,019)	77.4	76.5 (895)	76.9	0.80
Biological father alive	58.3 (1,853)	57.8	57.1 (953)	56.2	59.5 (900)	59.5	0.21
Always lives with biological father[e]	51.7 (944)	51.4	48.8 (458)	48.1	54.9 (486)	54.4	0.21
Sexual behaviour, % (*n*)							
Ever had sex	23.4 (754)	25.9	16.6 (280)	18.8	30.9 (474)	33.0	0.06
Age at first sex (median, IQR)[f]	16 (15–17)		17 (16–18)		16 (15–17)		
Total lifetime sex partners (median, IQR)[f]	2 (1–3)		1 (1–2)		2 (1–4)		
Total lifetime sex partners (mean; ± SD; range)[f]	2.5; ± 2.9, 0–30		1.4, ± 1.2, 1–17		3.2, ± 3.3, 0–30		
Partner relationship, % (*n*)[f]							
Younger or same age partner	59.9 (458)	59.0	34.8 (98)	32.4	74.5 (360)	74.4	< 0.001
1–4 years older	34.1 (261)	34.7	53.5 (151)	55.1	22.8 (110)	22.6	< 0.0001
≥ 5 years older	6.0 (46)	6.3	11.7 (33)	12.5	2.7 (13)	2.9	0.003
Pregnancy							
Prevalence			3.2 (52)	3.3			

[a]Missing values excluded from percentage calculation.
[b]Adjusted for school clusters.
[c]Adjusted for schools and comparing boys to girls.
[d]Proportion calculated for those reporting biological mother alive.
[e]Proportion calculated for those reporting biological father alive.
[f]Calculated for those reporting ever had sex.

ever having had sex, aOR = 2.8, 95% CI 1.9–4.2; $p < 0.001$. Although not significant, girls having a sex partner with an age difference at least 5 years and older were more likely to be HIV positive. HIV prevalence was higher in those 20 years and older (aOR = 4.2, 95% CI 1.3–2.6; $p = 0.01$) not having a living biological mother (aOR = 3.6, 95% CI 2.9–4.5; $p < 0.001$) and ever having had sex (aOR = 2.8, 95% CI 1.9–4.2; $p < 0.001$). There was no association between pregnancy and HIV (aOR = 0.6, 95% CI 0.1–2.2; $p = 0.39$). Of note, 21 (1.5%) girls who reported not ever having had sex tested positive for βHCG. In contrast, HIV infection in boys was higher in those without a living biological mother (aOR = 2.8, 95% CI 1.2–6.2; $p = 0.01$) or father (aOR = 2.0, 95% CI 1.3–3.2; $p = 0.002$).

Table 2. HIV-1 Prevalence And Demographic, Behavioral, and Biological Correlates Of Infection in High School Students in Rural Kwazulu-Natal, South Africa

	HIV Prevalence						
	Overall		*Girls*		*Boys*		
Characteristics	*Unadjusted*[a] n/N; %	*Adjusted*[a,b] % (95% CI)	*Unadjusted*[a] n/N; %	*Adjusted*[a,b] % (95% CI)	*Unadjusted*[a] n/N, %	*Adjusted*[a,b] % (95% CI)	p-*value*[c]
HIV prevalence	143/3241; 4.4	4.8 (2.9–6.7)	105/1698; 6.2	6.8 (3.9–9.8)	38/1542; 2.5	2.7 (1.6–3.8)	0.007
School specific							
School A	37/1024; 3.6	Not applicable	29/575; 5.0	Not applicable	8/448; 1.8	Not applicable	
School B	28/530; 5.3		22/286; 7.7		6/244; 2.5		
School C	22/335; 6.6		15/166; 9.0		7/169; 4.1		
School D	35/621; 5.6		26/299; 8.7		9/322; 2.8		
School E	21/731; 2.9		13/372; 3.5		8/359; 2.2		
Age specific							
12–15 years	50/1567; 3.2	3.8 (1.7–5.8)	33/882; 3.7	4.6 (1.9–7.3)	17/685; 2.5	2.7 (0.6–4.9)	0.17
16–17 years	31/991; 3.1	3.5 (0.8–6.2)	24/492; 4.9	5.3 (0.8–9.9)	7/499; 1.4	1.9 (1.2–2.6)	0.11
18–19 years	24/456; 5.3	4.9 (1.8–8.0)	19/206; 9.2	8.7 (0–17.7)	5/250; 2.0	3.4 (0–8.6)	0.28
≥20 years	38/223; 17.0	16.7 (8.3–25.2)	29/117; 24.8	23.1 (7.7–38.5)	9/106; 8.5	11.1 (2.7–19.4)	0.12
Grade specific							
Grade 8	33/732; 4.5	5.3 (1.4–9.2)	24/359; 6.7	8.5 (1.1–15.9)	9/373; 2.4	2.7 (0.9–4.6)	0.10
Grade 9	19/726; 2.6	2.7 (0.2–5.2)	12/369; 3.3	3.5 (1.7–5.4)	7/357; 2.0	3.1 (0–10.4)	0.79
Grade 10	28/791; 3.5	4.0 (0.2–7.8)	19/413; 4.6	5.6 (0.7–10.4)	9/378; 2.4	3.4 (0–7.4)	0.35
Grade 11	34/603; 5.6	6.0 (3.2–8.7)	26/334; 7.8	8.0 (4.6–11.3)	8/269; 3.0	4.0 (1.3–6.8)	0.04
Grade 12	29/383; 7.6	7.3 (4.2–10.3)	24/220; 10.9	10.9 (1.5–20.3)	5/163; 3.1	4.6 (1.2–8.1)	0.16
Family characteristics							
Biological mother alive							
Yes	72/2493; 2.9	3.0 (2.3–3.7)	53/1318; 4.0	4.3 (2.8–5.8)	19/1175; 1.6	1.6 (0.9–2.3)	0.002
No	71/743; 9.6	10.1 (4.7–15.5)	52/377; 13.8	14.4 (7.3–21.6)	19/366; 5.2	5.7 (1.3–10.1)	0.02
p value		0.02		0.02		0.06	
Always lives with biological mother[d]							
Always	52/1913; 2.7	2.7 (1.7–3.7)	38/1019; 3.7	3.8 (2.1–5.4)	14/894; 1.6(0.8–2.4)	1.6 (0.9–2.2)	0.01
Sometimes	17/474; 3.6	4.3 (1.7–6.9)	12/239; 5	6.5 (0.7–12.2)	5/235; 2.1(0.3–4)	2.4 (1.3–3.4)	0.12
Never	3/86; 3.5	5.6 (1.3–9.9)	3/46; 6.5	12.6 (0–33.7)	0/40	—	—
p value		0.09		0.05		0.08	
Biological father alive							
Yes	54/1852; 2.9	3.3 (1.4–5.2)	40/953; 4.2	4.8 (1.7–7.8)	14/899; 1.6	1.8 (0.8–2.8)	0.03
No	86/1327; 6.5	6.9 (4.3–9.6)	62/715; 8.7	9.3 (5.7–12.9)	24/612; 3.9	4.2 (2.5–5.9)	0.01
p value		0.01		0.03		0.01	
Always lives with biological father[e]							
Always	23/944; 2.4	2.7 (1.0–4.3)	18/458; 3.9	4.1 (1.5–6.7)	5/486;1.0	2.0 (0–5.0)	0.18
Sometimes	20/614; 3.3	3.9 (1.5–6.3)	15/338; 4.4	5.3 (1.3–9.3)	5/276; 1.8	2.7 (0.50–4.9)	0.18
Never	10/266; 3.8	4.0 (1.8–6.3)	6/143; 4.2	5.0 (2.0–7.9)	4/123; 3.3	5.3 (0–11.3)	0.88
p value		0.40		0.75		0.11	

(*continued*)

TABLE 2. (CONTINUED)

Characteristics	HIV Prevalence						
	Overall		*Girls*		*Boys*		
	Unadjusted[a] n/N; %	*Adjusted*[a,b] % (95% CI)	*Unadjusted*[a] n/N; %	*Adjusted*[a,b] % (95% CI)	*Unadjusted*[a] n/N, %	*Adjusted*[a,b] % (95% CI)	p-*value*[c]
Sexual behaviour							
Had sex							
Never	77/2471; 3.1	3.5 (1.6–5.4)	56/1411; 4.0	4.7 (1.8–7.5)	21/1060; 2.0	2.1 (0.9–3.2)	0.047
Ever	65/754; 8.6	8.8 (6.0–11.5)	48/280; 17.1	16.6 (6.4–26.7)	17/474; 3.6.0	4.8 (1.5–8.0)	0.03
p value		0.003		0.03		0.03	
Age at first sex[f]							
at ≤15 years	6/247; 2.4	3.8 (1.0–6.7)	4/58; 6.9	23.4 (0–89.3)	2/189; 1.1	2.9 (0–19.0)	0.06
at 16–17 years	27/335; 8.1	8.2 (4.2–12.3)	17/121; 14	17.0 (1.2–32.7)	10/214; 4.7	6.9 (0–14.6)	0.12
at 18–19 years	25/152; 16.4	15.6 (8.3–22.8)	21/83; 25.3	27.8 (20.9–34.7)	4/69; 5.8	11.8 (0–34.9)	0.01
at ≥20 years	7/34; 20.6	25.7 (0–54.4)	6/23; 26.1	51.9 (0–155.6)	1/11; 9.1	20.0	
p value		0.04		0.23		0.06	
Total lifetime sex partners[e]							
I only	33/379; 8.7	8.4 (5.3–11.5)	31/212; 14.6	14.1 (6.0–22.2)	2/167; 1.2	3.5 (0–14.8)	0.08
2 or more	32/387; 8.3	8.6 (5.2–12.1)	17/72; 23.6	35.0 (0.4–69.6)	15/315; 4.8	6.0 (2.6–9.4)	0.07
p value		0.89		0.15		0.21	
Partner relationship[e]							
Younger or same age partner	30/458; 6.6	6.8 (1.6–11.9)	15/98; 15.3	14.4 (2.6–26.2)	15/360; 4.2	5.6 (0.7–10.7)	0.07
1–4 years older	23/261; 8.8	8.6 (0.1–16.9)	22/151; 14.6	14.9 (0–31.0)	1/110; 0.9	3.0	—
≥5 years older	11/46; 23.9	24.6 (5.0–44.3)	10/33; 30.3	34.7 (4.9–64.5)	1/13; 7.7	50.0	—
p value		0.03		0.10		0.002	
Pregnancy status							
Not pregnant	97/1598; 6.1	6.8 (3.5–10.1)	97/1598; 6.1	6.8 (3.5–10.1)	Not applicable	Not applicable	—
Pregnant	5/52; 9.6	16.8 (0–122.6)	5/52; 9.6	16.8 (0–122.6)	Not applicable	Not applicable	
p value		0.44		0.44			

[a]Missing values excluded from percentage calculation.
[b]Adjusted for school clusters.
[c]Adjusted for schools and comparing boys to girls.
[d]Proportion calculated for those reporting biological mother alive.
[e]Proportion calculated for those reporting biological father alive.
[f]Calculated for those reporting ever had sex.

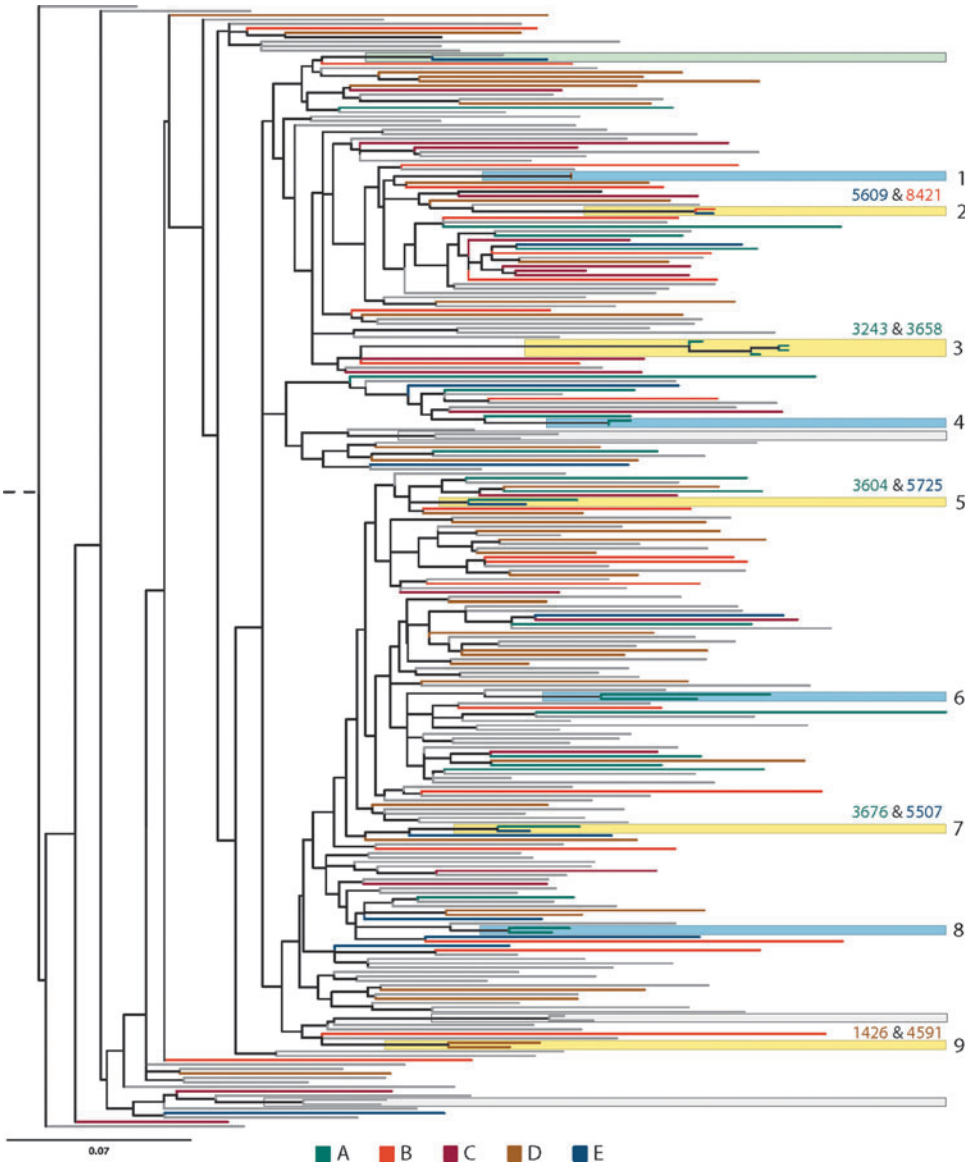

FIG. 1. Phylogenetic tree showing sequences derived from 120 students (106 from this study and 14 samples collected previously from Vulindlela) and 135 background sequences (Vulindlela and Durban). The extant branches of the phylogeny are color coded according to the school of origin **(A, B, C, D, E)** of the student or as background community sequence data (*gray branches*). The reconstructed phylogeny shows significant groupings (support values >90%) between two different students (*yellow boxes*), temporally separated sequences from the same individual (*blue boxes*), between a student and a background sequence (*green box*), and different individuals in the background sequence data (*gray boxes*). The nine significant groupings containing sequences from the same individual or two different students are numbered one through nine.

Discussion

In this survey, we found that one out of every 20 high school students was HIV infected, underscoring the magnitude of the epidemic in this rural setting. The gender disparity in HIV prevalence among girls and boys is striking, with our data highlighting an almost three times greater HIV risk in adolescent girls compared to their male peers. Though similar patterns of HIV prevalence have been reported in population-based surveys and from a representative sample of 15–24 year olds,[5] these have focused on out-of-school students. The high HIV prevalence in these adolescents also suggests that as yet there has been little or no impact of HIV prevention and treatment programs on HIV acquisition rates. A major concern is that HIV testing rates among adolescents remains low and the majority of HIV-infected individuals are unaware of their HIV status, posing a barrier to HIV prevention, care, and treatment efforts.[19] Moreover, as many of these recently infected adolescents are likely to be in the acute or early phases of HIV infection with potentially high viral loads, they also serve as reservoirs for HIV transmission sustaining the survival of the epidemic.

Over the past 20 years antenatal[1] and population-based[5,20,21] HIV surveillance has reasonably described the evolving epidemic in South Africa. However, as individuals benefit from increasing access to ART and the implementation of expanded options for HIV prevention, prevalence measures from current surveillance programs over time become less reliable. In adolescents, most HIV infections are relatively recent and HIV-related mortality is relatively low,[2–4] such that HIV prevalence in this age group will be useful to infer HIV incidence. Given that most young South Africans attend high school, these institutions may represent important and convenient venues for surveillance expansion to better assess trends in HIV infection. We found that undertaking such high school-based surveillance was feasible, with a student participation rate exceeding 80%.

In our preliminary phylogenetic analyses designed to yield clues as to where HIV transmission is occurring and/or being sustained, our analysis identified five clusters within schools,

Table 3. Association of Select Demographic, Behavioral, and Biological Characteristics and HIV in High School Students in Rural Kwazulu-Natal, South Africa

	Overall				*Girls*				*Boys*			
	OR (95% CI)	p *value*	*aOR*[a,b] *(95% CI)*	p *value*	*OR (95% CI)*	p *value*	*aOR*[b,c] *(95% CI)*	p *value*	*OR (95% CI)*	p *value*	*aOR*[a,b] *(95% CI)*	p *value*
Gender												
Boys	1.0		1.0									
Girls	2.6 (2.1–3.3)	<0.0001	3.0 (2.4–3.8)	<0.0001								
Age groups												
12–15 years	1.0		1.0		1.0		1.0		1.0		1.0	
16–17 years	0.98 (0.6–1.7)	0.94	0.8 (0.5–1.4)	0.46	1.3 (0.7–2.4)	0.36	1.1 (0.6–1.8)	0.83	0.6 (0.–1.0)	0.05	0.4 (0.2–0.9)	0.02
18–19 years	1.7 (1.1–2.5)	0.01	1.1 (0.8–1.5)	0.50	2.6 (1.3–5.3)	0.01	1.6 (0.9–2.8)	0.14	0.8 (0.2–3.2)	0.75	0.5 (0.2–1.7)	0.27
≥20 years	6.2 (2.9–13.6)	<0.0001	3.5 (1.4–8.9)	0.009	8.5 (3.3–21.6)	<0.0001	4.2 (1.3–2.6)	0.01	3.6 (1.2–11.1)	0.02	2.2 (0.8–6.3)	0.15
Grade[d]												
Grade 8	1.0				1.0				1.0			
Grade 9	0.6 (0.4–0.9)	0.01			0.5 (0.3–0.7)	0.001			0.8 (0.4–1.8)	0.59		
Grade 10	0.8 (0.3–1.8)	0.56			0.7 (0.2–2.1)	0.50			1.0 (0.4–2.5)	0.98		
Grade 11	1.3 (0.6–2.6)	0.52			1.2 (0.5–2.7)	0.70			1.2 (0.4–3.9)	0.71		
Grade 12	1.7 (0.97–3.1)	0.06			1.7 (0.8–3.9)	0.20			1.3 (0.3–4.8)	0.71		
Biological mother alive												
Yes	1.0		1.0				1.0				1.0	
No	3.6 (2.6–4.8)	<0.0001	3.3 (2.4–4.5)	<0.0001	3.8 (3.2–4.6)	<0.0001	3.6 (2.9–4.5)	<0.001	3.3 (1.6–6.7)	<0.001	2.8 (1.2–6.2)	0.01
Lives with biological mother												
Always	1.0				1.0				1.0			
Sometimes	1.3 (0.9–2.0)	0.14			1.4 (0.8–2.4)	0.28			1.4 (1.1–1.8)	0.02		
Never	1.3 (0.6–2.7)	0.48			1.8 (0.8–3.9)	0.14			—	—		
Biological father alive												
Yes	1.0		1.0		1.0		1.0		1.0		1.0	
No	2.3 (1.5–3.4)	<0.0001	1.6 (1.0–2.6)	0.05	2.2 (1.4–3.4)	<0.001	1.5 (0.9–2.6)	0.17	2.6 (1.8–3.7)	<0.0001	2.0 (1.3–3.2)	0.002
Lives with biological father												
Always	1.0				1.0				1.0			
Sometimes	1.3 (1.1–1.7)	0.01			1.1 (0.7–1.9)	0.63			1.8 (0.3–10.0)	0.51		
Never	1.6 (1.3–1.9)	<0.0001			1.1 (0.5–2.3)	0.86			3.2 (0.5–20.8)	0.22		
Had sex												
Never	1.0		1.0		1.0		1.0		1.0		1.0	
Ever	2.9 (2.0–4.4)	<0.0001	2.3 (1.8–3.0)	<0.0001	5.0 (2.7–9.2)	<0.0001	2.8 (1.9–4.2)	<0.001	1.8 (0.7–5.1)	0.24	1.7 (0.7–4.1)	0.21
Total lifetime sex partners[e]												
I only	1.0				1.0				1.0			
2 or more	0.9 (0.6–1.4)	0.78			1.7 (0.9–3.3)	0.08			5.0 (1.7–10.0)	0.001		
Partner relationship[e]												
Younger or same age partner	1.0				1.0				1.0			
1–4 years older	1.4 (0.4–4.6)	0.60			0.9 (0.3–3.2)	0.92			0.2 (0–2.1)	0.18		
≥5 years older	4.5 (1.5–13.6)	0.01			2.4 (0.7–8.4)	0.17			1.9 (0.2–16.9)	0.56		
Pregnant												
No	1.0				1.0		1.0					
Yes	1.6 (0.3–8.1)	0.54			1.6 (0.3–8.1)	0.54	0.6 (0.1–2.2)	0.39				

[a]Adjusted odds ratio (aOR) from multivariable model adjusting for gender, age, biological mother alive, biological father alive, and had sex.
[b]Adjusted for school clusters ($n = 5$).
[c]Adjusted odds ratio (aOR) from multivariable model adjusting for gender, age, biological mother alive, biological father alive, had sex, and pregnancy.
[d]Grade not included in the final multivariable model as it highly correlated with age groups.
[e]Not included in the multivariable model due to small sample sizes.

which suggests that students are observed as groupings in a highly supported putative transmission cluster when compared to background data. Although HIV infection was not associated with intraschool transmission in this rural setting, the source of infection and the direction of HIV transmission, although difficult to ascertain, require greater coverage and further analysis of schools and community sequences. To better understand HIV transmission dynamics in adolescents, larger sample sizes with robust study designs, which include more diverse populations, are needed to understand viral diversity for targeting of interventions. These studies will be critical, as any reliable data on infection dynamics from southern African settings are limited.[22] As biomedical interventions are designed and rolled-out, it will be vital to understand the proportion of HIV infections that may be attributed to any major source of transmission over time. Such understanding requires improved geographically representative surveillance with phylogenetic and epidemiological analyses.[23]

We report that at least one in four students was sexually active; sexual debut occurred at a young age of around 15 years with a mean of about two lifetime sex partners. Many adolescents in comparably high HIV burden settings also have a high burden of sexually transmitted infections,[24] pregnancies,[7,25,26] alcohol use,[27] and substance use.[28] Although we found no association between pregnancy and HIV, HIV prevalence was higher in pregnant girls, and our data underscore the importance of linking HIV prevention efforts to a broader SRH and healthy lifestyle education. Certainly, the observed prevalence of HIV infection in those female students <20 years of age who engage with sex partners at least 5 years their senior should be incorporated as a key message within school HIV prevention programs, and in the longer term age mixing within grades may need to be addressed within South African schools.

Previous literature has indicated that cohesive family structures are important determinants of risk, as parents could help instill and reinforce messages of protection from risky behavior.[29] Our data confirm that students who have lost parents are at increased risk of HIV, and highlight that identification of these young at-risk populations is critical as we identify priority groups for HIV prevention interventions.

A key strength in our study is the large sample size with a response rate of over 80%; however, there are several limitations. Being a cross-sectional study, we were unable to establish the temporality of risk behaviors and HIV infection; we hope that future surveillance will enable such information to be inferred. Furthermore, we report some social desirability bias, with 1.5% of girls who reported never having had sex testing positive for βHCG. Although the phylogenetic analyses suggest that infection among students is not associated with intraschool transmission in this rural South African setting, larger sexual transmission network studies are needed to better understand HIV transmission dynamics. However, in spite of these limitations this study provides valuable information on the burden of HIV infection and prevalent risk behaviors, and provides evidence that the robust and adolescent-focused surveillance systems required to measure new infections in a mature epidemic setting are feasible.

In conclusion, the epidemiology of HIV infection in this community is dynamic and complex. The high burden of infection among students, the first phylogenetic study among high school students showing limited intraschool transmission, and the absence of a living parent underscore the importance of encouraging schools to intensify HIV surveillance programs in order to monitor the epidemic, identify transmission patterns, and provide comprehensive HIV prevention packages to reduce adolescent's vulnerability to HIV infection.

Acknowledgments

We are grateful to all the high school students and the study staff for making this study possible. A special thanks to Reshoketswe Matlala, Fanelesibonge Ntombela, parents and guardians of students, principals, educators, schools governing boards' members, and the CAPRISA Community Research Support Groups. This study would not have been possible without the support of the KwaZulu-Natal departments of Education and Health. Sincere thanks to Dr. Stephen Korsman for assistance with the optimization of DBS laboratory work, Debbie Stewart for finalizing sequences and the administration of the sequencing project, Melissa-Rose Abrahams for assistance with the optimization of DBS laboratory work and critical reading of sections of the manuscript, and Johanna Ledwaba for extraction, amplification, and chromatogram assembly.

CAPRISA was established as part of the Comprehensive International Program of Research on AIDS (CIPRA) and supported by the National Institute of Allergy and Infectious Disease (NIAID), National Institutes of Health (NIH) and the U.S. Department of Health and Human Services (DHHS) (Grant 1 U19 AI51794). The U.S. President's Emergency Plan for AIDS Relief (PEPFAR) Strategic Information grant supported the HCT programme. A.B.M.K. and J.A.F. career development was supported by Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP) funded by the Fogarty International Center, National Institutes of Health (Grant D43TW00231) for training support. T.B. was supported by the Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP), Implementation Science Scholarship Program funded by the United States President's Emergency Plan for AIDS Relief (PEPFAR) through the Fogarty International Center, National Institutes of Health (Grant D43TW00231).

A.B.M.K., Q.A.K., and S.S.A.K. conceptualized the project and were responsible for the protocol development, survey design, analysis, interpretation of results, and manuscript writing. A.B.M.K., T.B., G.M., and J.A.F. were responsible for field implementation of the project, data collection, and critically reviewed the final manuscript. N.S. was responsible for the laboratory assessments and reviewed the manuscript. N.Y.Z. provided statistical support and critically reviewed the manuscript. C.W., S.T., and J.C.M. contributed to the phylogenetic sequencing protocol, analysis, and interpretation. R.D. critically reviewed the manuscript. All authors have read and approved the final manuscript.

Author Disclosure Statement

No competing financial interests exist.

References

1. South Africa Department of Health: The 2011 National Antenatal Sentinel HIV & Syphilis Prevalence Survey in South Africa, 2012. www.health.gov.za/docs/reports/2013/Antenatal_survey_report_2012_web_optimized.pdf.
2. Ghys PD, Kufa E, and George MV: Measuring trends in prevalence and incidence of HIV infection in countries with generalised epidemics. Sex Transm Infect 2006;82(Suppl 1): i52–56.
3. Gouws E, Stanecki KA, Lyerla R, and Ghys PD: The epidemiology of HIV infection among young people aged 15–24 years in southern Africa. AIDS 2008;22(Suppl 4):S5–16.
4. Mahy M, Chhea C, Saliuk T, Varetska O, and Lyerla R: A proxy measure for HIV incidence among populations at increased risk to HIV. J HIV/AIDS Surveill Epidemiol 2010;2(1):8.
5. Pettifor AE, Rees HV, Kleinschmidt I, *et al.:* Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. AIDS 2005;19(14):1525–1534.
6. Lewis F, Hughes GJ, Rambaut A, Pozniak A, and Leigh Brown AJ: Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med 2008;5(3):e50.
7. Kharsany AB, Mlotshwa M, Frohlich JA, *et al.:* HIV prevalence among high school learners—opportunities for schools-based HIV testing programmes and sexual reproductive health services. BMC Public Health 2012;12:231.
8. Middelkoop K, Rademeyer C, Brown BB, *et al.:* Epidemiology of HIV-1 subtypes among men who have sex with men in Cape Town, South Africa. J Acquir Immune Defic Syndr 2014;65(4):473–480.
9. Tatt ID, Barlow KL, and Clewley JP: A gag gene heteroduplex mobility assay for subtyping HIV-1. J Virol Methods 2000;87(1–2):41–51.
10. Schultz AK, Zhang M, Leitner T, *et al.:* A jumping profile hidden Markov model and applications to recombination sites in HIV and HCV genomes. BMC Bioinform 2006; 7:265.
11. de Oliveira T, Deforche K, Cassol S, *et al.:* An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 2005;21(19):3797–3800.
12. Alcantara LCJ, Cassol S, Libin P, *et al.:* A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences. Nucl Acids Res 2009;37(Suppl 2):W634–W642.
13. Wright IA and Travers SA: RAMICS: Trainable, high-speed and biologically relevant alignment of high-throughput sequencing reads to coding DNA. Nucl Acids Res 2014; 10.1093/nar/gku473.
14. Chopera DR, Mann JK, Mwimanzi P, *et al.:* No evidence for selection of HIV-1 with enhanced gag-protease or Nef function among breakthrough infections in the CAPRISA 004 tenofovir microbicide trial. PLoS One 2013;8(8):e71758.
15. Chopera DR, Woodman Z, Mlisana K, *et al.:* Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS Pathog 2008; 4(3):e1000033.
16. Ntale RS, Chopera DR, Ngandu NK, *et al.:* Temporal association of HLA-B*81:01- and HLA-B*39:10-mediated HIV-1 p24 sequence evolution with disease progression. J Virol 2012;86(22):12013–12024.
17. Price MN, Dehal PS,and Arkin AP: FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26(7):1641–1650.
18. Stamatakis A: RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22(21):2688–2690.
19. Binagwaho A, Nutt CT, Nsanzimana S, Wagner CM, and Mukherjee JS: Children and adolescents with HIV. Lancet Infect Dis 2013;13(8):654.
20. Shisana O, Rehle T, Simbayi LC, *et al.: South African National HIV Prevalence, Incidence, Behaviour and Communication Survey 2008: A Turning Tide Among Teenagers*? HSRC Press, Cape Town, 2009.
21. Shisana O, Rehle T, Simbayi LC, *et al.: South African National HIV Prevalence, Incidence and Behaviour Survey, 2012.* HSRC Press, Cape Town, 2014.
22. Wertheim JO, Leigh Brown AJ, Hepler NL, *et al.:* The global transmission network of HIV-1. J Infect Dis 2014; 209(2):304–313.
23. Grabowski MK and Redd AD: Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS 2014;9(2):126–133.
24. Jewkes R, Nduna M, Levin J, *et al.:* Impact of stepping stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: Cluster randomised controlled trial. BMJ 2008;337:a506.
25. Abdool Karim SS, Churchyard GJ, Abdool Karim Q, and Lawn SD: HIV infection and tuberculosis in South Africa: An urgent need to escalate the public health response. Lancet 2009;374(9693):921–933.
26. Panday S, Makiwane M, Ranchod C, and Letsoala T: Teenage pregnancy in South Africa: With a specific focus on school-going learners. Human Sciences Research Council, 2009; www.hsrc.ac.za/Research_Publication-21277.phtml. Accessed October 2013.
27. Onya H, Tessera A, Myers B, and Flisher A: Adolescent alcohol use in rural South African high schools. Afr J Psychiatry 2012;15:352–357.
28. Taylor M, Jinabhai CC, Naidoo K, Kleinschmidt I, and Dlamini SB: An epidemiological perspective of substance use among high school pupils in rural KwaZulu-Natal. S Afr Med J 2003;93:136–140.
29. DiClemente RJ, Wingood GM, Crosby R, *et al.:* Parental monitoring: Association with adolescents' risk behaviors. Pediatrics 2001;107(6):1363–1368.

Address correspondence to:
Ayesha B.M. Kharsany
CAPRISA
Doris Duke Medical Research Institute, 2nd Floor
Nelson R. Mandela School of Medicine
University of KwaZulu-Natal/Private Bag 7
Congella
Durban 4013
South Africa

E-mail: kharsany@ukzn.ac.za