Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2008:**37**;88–105

doi:10.1093/ije/dym255

HIV incidence in 3 years of follow-up of a Zimbabwe cohort—1998–2000 to 2001–03: contributions of proximate and underlying determinants to transmission

Ben Lopman,[1] Constance Nyamukapa,[1,2] Phyllis Mushati,[2] Zivai Mupambireyi,[2] Peter Mason,[2] Geoff P Garnett[1] and Simon Gregson[1,2]*

Accepted 29 November 2007

Background In recent years, HIV prevalence has begun to decline in Zimbabwe, which has been associated with reductions in sexual risk behaviour. Here, we analyse the determinants of HIV incidence in this period of decline and estimate the population-level impact of identified risk factors.

Methods A population-based cohort of 1672 HIV-negative adult males and 2465 HIV-negative adult females was recruited between 1998 and 2000. Each individual was then followed-up 3 years later. The influence and inter-relationship of social, behavioural and demographic variables were examined using a proximate determinants framework. To explore the population-level influence of a variable, methods were developed for estimating a risk factor's contribution to the reproductive number (CRN).

Results HIV incidence was 19.9 [95% confidence interval (CI) 16.3–24.2] per 1000 person years in men and 15.7 (95% CI 13.0–18.9) in women. Multiple sexual partners, having an unwell partner, and reporting another sexually transmitted disease were risk factors that captured the main aspects of the proximate determinants framework: individual behaviour, partnership characteristics and the probability of transmission, respectively. If the proximate determinants fully captured risk of HIV infection, underlying factors would not influence a fully parameterized model. However, a number of underlying social and demographic determinants remained important in regression models after including the proximate determinants. For both sexes, having multiple sexual partners made a substantial CRN, but, for women, no behaviour explained more than 10% of new infections.

[1] Department of Infectious Disease Epidemiology, Imperial College London, UK.

[2] Biomedical Research and Training Institute, Harare, Zimbabwe.

* Corresponding author. Department of Infectious Disease Epidemiology, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK.
E-mail: sajgregson@aol.com

Conclusions The proximate determinants did not explain the majority of new infections at the population level. This may be because we have been unable to measure some risks, but identifying risk factors assumes that those acquiring infections are somehow different from others who do not acquire infections. That they are not suggests that in this generalized epidemic there is little difference in readily identifiable characteristics of the individual between those who acquire infection and those who do not.

Keywords HIV/AIDS, epidemiological methods, Zimbabwe, incidence, population attributable fraction, Africa

Introduction

Despite the wide spread of HIV in sub-Saharan Africa relatively few studies have followed cohorts of individuals exploring risk factors for acquiring infection. Published examples include cohorts from Uganda, Tanzania and Zimbabawe.[1–4] The sequence of risks and outcomes only becomes transparent if we observe changes amongst individuals and populations over time in cohort studies. Such studies are critically important to the design and targeting of prevention activities. They also allow us to explore changing patterns of acquisition as the HIV epidemic spreads and how the community responds to it, but the wide range of risks, the exploratory nature of many statistical analyses and the use of epidemiological methods developed to explore the aetiology of chronic rather than infectious disease limits our ability to interpret results. Here we analyse cohort data from rural Zimbabwe exploring both patterns of risk following declines in prevalence and also the methods available to understand these risks.

Zimbabwe is experiencing a widespread HIV epidemic with prevalence having peaked at around 25% in the adult population in the late 1990s.[5] However, recent evidence suggests that HIV prevalence is now declining in Zimbabwe[6,7] as it may be in a few other countries with generalized epidemics, such as Haiti and Kenya.[8] The most recent nationally representative estimate of prevalence in Zimbabwe was 18%.[9] Prevalence and incidence may rise and then fall as a 'natural' pattern for an epidemic but in Zimbabwe the prevalence decline has been associated with observed reductions in casual partnerships, a delay in sexual debut and increases in condom use.[7]

Epidemics of sexually transmitted diseases may be conceptualized as progressing through phases.[10] Due to the long incubation of AIDS, which ultimately results in death, the effect of interventions or behaviour change on prevalence will be slow to accrue and difficult to observe empirically.[8] As the HIV epidemic passes through different phases we can expect that the risks and the context of those acquiring infection will change. Therefore, this recent period in Zimbabwe, when prevalence began to decline, is a particularly important phase in which to analyse and understand risk factors for incidence.

The proximate determinants framework of HIV transmission elaborated by Boerma and Weir provides a conceptual basis for examining both underlying social and demographic determinants as well as individual level proximate and biological determinants predicting risk of HIV infection.[11] Within this framework, the proximate determinants can be affected by contextual factors and intervention programs. The proximate determinants (such as condom use)—which can be measured in empirical studies—are markers for biological determinants (such as transmission probability per contact) which are unobserved but ultimately lead to transmission. In the framework, underlying determinants act through the proximate determinants, and ultimately through biological determinants. For example, poverty may predispose females to sex work, which exposes them to high number of sexual partners and, thereby, to greater exposure to infection. Previously we have explored risk factors associated with prevalent HIV infections from our baseline survey. Analyses of incidence are needed to understand the causal process leading to new infections as well as to disentangle historical from contemporary trends.

There is no single correct or straightforward statistical method for analysing data using this framework. We take two approaches: first, we explore the ability of measured proximate determinants to explain the influence of underlying determinants; then, second, we attempt to evaluate the importance of each determinant of infection in terms of how it influences the mean number of secondary cases arising from a new case (the effective reproductive ratio or R).[8,12]

Methods

Population

The Manicaland HIV/STD Prevention Project is an ongoing population-based open cohort study, full details of which can be found elsewhere.[6] The study population were resident in four subsistence farming areas, two roadside trading centres, four forestry, tea and coffee estates and two small towns in the rural

province of Manicaland in eastern Zimbabwe. All local residents were enumerated in an initial household census (conducted between July 1998 and February 2000) which was repeated 3 years later in each site. Males aged 17–54 years and females aged 15–44 years were recruited into a cohort study of HIV transmission. Every household in the study area was visited up to three times or until interviews were completed. To be eligible for the cohort, individuals had to be regular members of the household. Only one member of a marital pair was recruited.

At baseline and follow-up, demographic, socioeconomic and sexual behaviour data were collected through an interviewer-led questionnaire. To reduce social desirability bias, responses to sensitive questions about sexual behaviour were collected using an informal confidential system, in which responses were written on ballot slips and placed in a locked box.[6] Dried blood spots were collected for HIV serological testing for the purposes of the research only which was performed using a highly sensitive and specific antibody dipstick assay.[13] Different tests were used in baseline and follow-up (ICL-HIV 1 & 2 Dipstick, Thailand (baseline) and Abbott 3rd Generation HIV 1 & 2 EIA, USA or Genelavia MIXT HIV1&2, Sanofi Diagnostics Pasteur S.A., France (follow-up), which have been shown to work equivalently in Zimbabwe.[6] As was standard practice at the time the data collection was undertaken, participants were not given test results but were given a voucher for counselling and testing and were provided with the services of a mobile clinic (details are published elsewhere[6]). Lifetime uptake increased from 6% to 11% from baseline to follow-up; 21.5% of those who accepted testing were HIV-positive. The majority who went for testing reported either a pre- or post-test counselling session that included information on preventing infection. Anti-retrovirals were not available at the time of study, though efforts have been made such that the study population is now being included in the first wave of anti-retroviral programmes in Zimbabwe. Local clinics provided treatment of opportunistic infections, though availability was reported to be variable during the study period. In collaboration with local NGOs, the study nurses monitored and supplemented cotrimoxazole stocks at local clinics. Written informed consent was obtained as a condition of enrolment and continuation in the study. Participants unable to sign their name provided a finger print after being explained about the study by research assistants. Prior ethical approval for the study was obtained from the Research Council of Zimbabwe—Number 02187—and the Applied and Qualitative Research Ethics Committee in Oxford, United Kingdom—N97.039.

Follow-up

As reported previously 54% (2242/4142) of the males and 66% (3265/4922) of the females interviewed at baseline—and not known to have died subsequently—were re-interviewed at follow-up.[6] Follow-up rates of individuals were lower within households of higher wealth status (chi-squared $P < 0.0001$), better education (primary/none: 70%; secondary/higher: 55%, chi-squared $P < 0.001$) and being more mobile at baseline (64%, 53%, chi-squared $P < 0.0001$). Amongst the follow-ups, we analysed individuals at risk of seroconversion in HIV negative males (1777/2242) and females (2566/3265). Individuals with missing key data or ambiguous HIV test results were dropped leaving 1672 (94%) males and 2465 (96%) females for analysis. Date of seroconversion was randomly assigned (uniform distribution) between the last negative and first positive test. Follow-up time was censored at this date.

Statistical methods

Table 1 documents the variables that were investigated and whether they were modelled as proximate or underlying determinants. The current analysis is based upon the risk factors reported at follow up where individuals described both recent behaviours and those occurring in the period following their first interview. The association between the determinants and HIV incidence was investigated by fitting Poisson regression models. Variables (or stratum of variables) determined to improve crude models or age adjusted models (Wald test P-value $\leqslant 0.05$) are presented and were retained for multivariable analysis. When a set of variables was highly correlated (correlation coefficient >0.5) and measured a similar behaviour, the variable most highly associated with HIV incidence was retained for the multivariable model. The role of a variable was assessed by comparing a model without the variable to a full model using the likelihood ratio test.

Separate models were constructed for proximate determinants and underlying determinants.[14] Then, significant proximate determinants were added to the underlying determinants model. If the underlying determinants all acted through the proximate determinant, the underlying determinants should no longer substantially influence a model containing terms for the proximate determinants.

Population-level impact of risk factors

Because of the dynamics of how an infectious disease moves through a population, there is no standard statistical method to decompose the importance of particular determinant on the totality of transmission, such as the population attributable fraction (PAF) for non-infectious diseases.[10,11,15,16] We propose that determinants of infection can be conceptualized as having a bearing on the reproductive ratio, R, by influencing its constituent components of: (i) the rate of exposure to infectious persons; (ii) the transmission probability; and (iii) the duration of infectiousness. Ideally, we would be able to track the spread of infection from individual to individual and examine the impact of risk in both the infective and

Table 1 Variables modelled as proximate or underlying determinants

Biological determinants	Potential proximate determinants	Potential underlying determinants
Exposure of susceptible to infected persons	Number of partners during risk period Frequency of intercourse Concurrent partners HIV prevalence among opposite sex in community Years sexually active during follow-up Extramarital sex	Age group Marital status Religion Education Work status Socioeconomic status Community type Mobility/migration
Per contact transmission probability	Partnership characteristics: Partners health Living arrangements Partner has other partners Age of partner Age difference Regular or casual partnership Met partner here or away Condom use Practice of dry sex Sexually transmitted infections Male circumcision	Beer hall visits Paying/being paid for sex Previous HIV test Pregnancy Knowing/caring for AIDS patient Belief and attitudes on: acceptability of condoms self-perceived risk of AIDS role of married men beer drinking Knowledge index of HIV
Duration of infectivity	None	

susceptible partners. However, whilst we have data on the infected and those who infect them, we do not have information on the links between them. We approximate R at time t (R_t) for the unexposed and exposed population of men and women using an incidence: prevalence ratio as described previously[6]:

$$R_t \approx \left\{\frac{G(t,t+n)}{n}\right\}\frac{D}{F(t)} \quad (1)$$

where $G(t, t+n)$ is the number of new HIV infections occurring between time t and time $t+n$, D is the duration of infectiousness in an infected person, and $F(t)$ is the number of infected people with HIV in the population at baseline including those who die by time $t+n$. n is the duration of follow-up (3 years in this case) and D is the duration of infectiousness (assumed to be 10 years, the mean time from infection to death in this population, where at the time access to antiretroviral treatment was extremely limited).

We then estimate the reproductive number at time t for a hypothetical cohort where the risk factor was absent (R_{tu}), as described in the Appendix. The contribution to the reproductive number (CRN) is estimated as one minus the ratio R_{tu} in the unexposed sub-cohort to R_t in the entire cohort, using the formula

$$CRN \approx \left[\frac{1-R_{tu}}{R_t}\right] \quad (2)$$

This method provides an imperfect measure since it assumes that all transmission stays within the group exposed to a particular risk. Therefore, infections occurring across groups (i.e. from high risk to low risk groups) do not contribute to the estimate of R. The true value of R is affected by the behaviour of both the infected and susceptible population, although only the susceptible population is considered in this method. Therefore the CRN is a conservative underestimate of the real influence of a risk factor (see the Appendix). The CRN does—unlike the PAF—take some account of transmission as well as acquisition and can account for the changing size of the exposed populations.

Results

Incidence of HIV

Ninety-eight men seroconverted in a total of 4916 years of follow-up, giving an incidence rate (IR) of 19.9 [95% confidence interval (CI) 16.3–24.2] per 1000 person years. One hundred and thirteen incident cases amongst females in 7184 years of follow-up resulted in a marginally lower IR of 15.7 (95% CI 13.0–18.9). Age-standardizing the rates made little difference (male 20.4, female 15.4) compared with the crude rates.

Proximate determinants

Tables 2 and 3 show the proximate determinants associated with HIV incidence in men and women, respectively. For men, in multivariable models controlling for other proximate determinants, multiple partners ($RR = 2.4$), experience of genital sores ($RR = 2.7$), having an unwell partner ($RR = 1.8$), and local prevalence amongst women ($RR = 1.3$, per 10%

Table 2 Proximate determinants of incident HIV infection: Males

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Adjusted[b] RR
Total regular partners since baseline (last 3 years)				
0	18/1622	11.2	1**	
1	46/2251	20.3	1.8 (1.1–3.1)*	
Multiple	32/1035	31.1	2.8 (1.6–5)**	
Total non-regular partners since baseline (last 3 years)				
0	46/2741	16.7	1	
1	20/838	24	1.4 (0.8–2.4)	
Multiple	30/1316	22.8	1.4 (0.9–2.2)	
Total partners since baseline (last 3 years)				
0	8/902	8.9	1***	1*
1	27/1743	15.4	1.7 (0.8–3.8)	1.4 (0.6–3.1)
Multiple	61/2251	27.2	3.1 (1.5–6.4)**	2.4 (1.1–5.0)*
Sex outside marriage[c]				
No	31/2217	13.5	1	
Yes	67/2690	24.0	1.8(1.2–2.8)**	
Partners in last month				
0	32/1965	16.2	1	
1	53/2552	20.7	1.3 (0.8–2)	
Multiple	13/398	33.0	2.0 (1.1–3.9)*	
Current sexual partners[d]				
0	19/1403	13.4	1*	1
1	63/3021	20.7	1.5 (0.9–2.6)	1.2 (0.7–2.1)
Multiple	16/510	32.5	2.4 (1.2–4.7)**	1.5 (0.7 – 3.0)
Experienced genital pain or discharge in last year[e]				
No	85/4593	18.3	1	
Yes	13/322	40.4	2.2 (1.2–3.9)**	
Experienced genital sores in last year[e]				
No	84/4692	17.9	1	1
Yes	14/224	61.5	3.4 (2–6.1)***	2.7 (1.5–4.8)**
HIV prevalence in other sex in community at baseline				
Per 10%			1.3 (1–1.6)*	1.2 (1.0–1.5)*
Partner's health				
Well	82/4585	17.7	1	1
Unwell	16/331	44.8	2.5 (1.5–4.4)**	1.8 (1.1–3.0)**
Consistent condom use in regular partnership				
No	53/2117	25.0	1	
Yes	3/243	12.3	0.5 (0.2–1.5)	
Thinks that spouse has other sexual partners				
No	85/3986	21.3	1	
Yes	7/366	19.1	0.9 (0.4–1.9)	

(continued)

Table 2 Continued

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Adjusted[b] RR
Circumcised				
No	92/4634	19.8	1	
Yes	6/314	19.1	0.9 (0.4–2.1)	
Practice dry sex				
No	77/3867	19.9	1	
Yes	2/162	12.3	0.5 (0.1–2.2)	
Last partner				
Regular	59/2543	23.1	1	
Casual	7/373	18.8	0.8 (0.4–1.8)	
Consistent condom use with non-regular partner[c]				
No	5/175	28.4	1	
Yes	3/206	3/206	0.5 (0.1–2.1)	

Other variables ($P > 0.05$) included: number of sex acts (linear variable), partner's age (categorized <25, 25 to 34, >34). Stars in the reference group (e.g. zero total partners since baseline) signifies *P*-value of including all strata of the variable in a model compared with a model without the variable.

[a]Seroconversions per 1000 person-years.

[b]Adjusted in Poisson regression models for other identified proximate determinants. Highly correlated variables were removed from multivariable models (total regular partners and total non-regular partners and sex outside of marriage with total partners: correlation coefficient (cc) = 0.62, 0.71, 0.81, respectively; partners in last month with current sexual partners: cc = 0.65; genital pain/discharge with genital sores: cc = 0.22).

[c]Variable constructed and defined as (i) all unmarried individuals who had a regular or non-regular partner or (ii) married indiviudals who had more than one regular partner or any non-regular partners in follow-up period.

[d]Number of sexual partners respondent believed he was involved with in around the time of interview.

[e]Self reported.

$^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$

Table 3 Proximate determinants of incident HIV infection: females

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Adjusted[b] RR
Total regular partners since baseline (last 3 years)				
0	23/1676	14.2	1***	
1	73/5257	14.2	1 (0.6–1.6)	
Multiple	17/251	67.2	4.7 (2.5–8.9)***	
Total non-regular partners since baseline (last 3 years)				
0	96/6697	14.7	1***	
1	6/371	16.2	1.1 (0.5–2.5)	
Multiple	11/115	94.4	6.4 (3.4–12)***	
Total partners since baseline (last 3 years)				
0	19/1510	13.1	1***	1
1	69/5223	13.5	1 (0.6–1.7)	1.0 (0.6–1.8)
Multiple	25/450	55.4	4.2 (2.3–7.7)***	3.2 (1.7 – 6.3)**
Last partner				
Regular	56/4258	13.4	1	
Casual	9/117	50.2	3.7 (1.8 – 7.5)***	
Consistent condom use with regular partner[c]				
No	42/3287	11.0	1	1
Yes	7/211	43.9	2.7 (1.2–6)*	1.0 (0.9–1.1)

(continued)

Table 3 Continued

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Adjusted[b] RR
Consistent condom use with non-regular partner[c]				
No	2/86	0	1	
Yes	4/42	44.1	4.1 (0.8–22.6)	
Sex outside marriage[d]				
No	74/6089	12.4	1	
Yes	39/1095	35.7	2.8 (1.9–4.2)***	
Partners in last month				
0	48/3157	15.7	1***	1**
1	58/3997	14.7	0.9 (0.6–1.4)	1.1 (0.7–1.8)
2+	7/271	244	15.6 (7–34.4)***	8.8 (3.5 – 21.1)***
Current partners[e]				
0	29/2021	14.8	1*	
1	80/5104	15.9	1.1 (0.7–1.6)	
2+	4/56	69.6	4.7 (1.7–13.4)**	
Experienced genital pain or discharge in last year[f]				
No	84/6032	14.2	1	
Yes	29/1151	25.7	1.8 (1.2–2.8)**	
Experienced genital sores in last year[f]				
No	99/6785	14.9	1	1
Yes	14/398	35.4	2.4 (1.4–4.1)***	2.0 (1.2–3.6)**
Ill in the last month				
No	66/4918	13.6	1	
Yes	47/2257	21.1	1.6 (1.1–2.3)*	
Partner's health				
Well	91/6249	14.8	1	1
Unwell	22/934	24.3	1.6 (1.1–2.6)*	2.0 (1.1–2.8)*
Practice dry sex				
No	88/4969	17.7	1	
Yes	22/1825	12.1	0.7 (0.4–1.1)	
HIV prevalence in other sex in community at baseline				
Per 10%			1.0 (0.7–1.5)	
Thinks that spouse has other sexual partners				
No	62/4174	14.8	1	
Yes	33/1491	22.1	1.5 (1.0–2.3)*	1.1 (0.7–1.9)

Stars in the reference group (e.g. partners in last month) signifies *P*-value of including all strata of the variable in a model compared with a model without the variable.

[a]Seroconversions per 1000 person-years.

[b]Adjusted in Poisson regression models for other identified proximate determinants.

[c]Consistent condom use defined as reporting using a condom throughout every sex act in the last two weeks. Analysis was based on two most recent partnerships. Men often report their spouse as their most recent partner and sometimes a casual partner as previous.

[d]Variable constructed and defined as (**i**) all unmarried individuals who had a regular or non-regular partner or (**ii**) ''married individuals who had more than one regular partner or any non-regular partners in follow-up period.

[e]Number of sexual partners respondent believed he was involved in around the time of interview.

[f]Self reported.

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

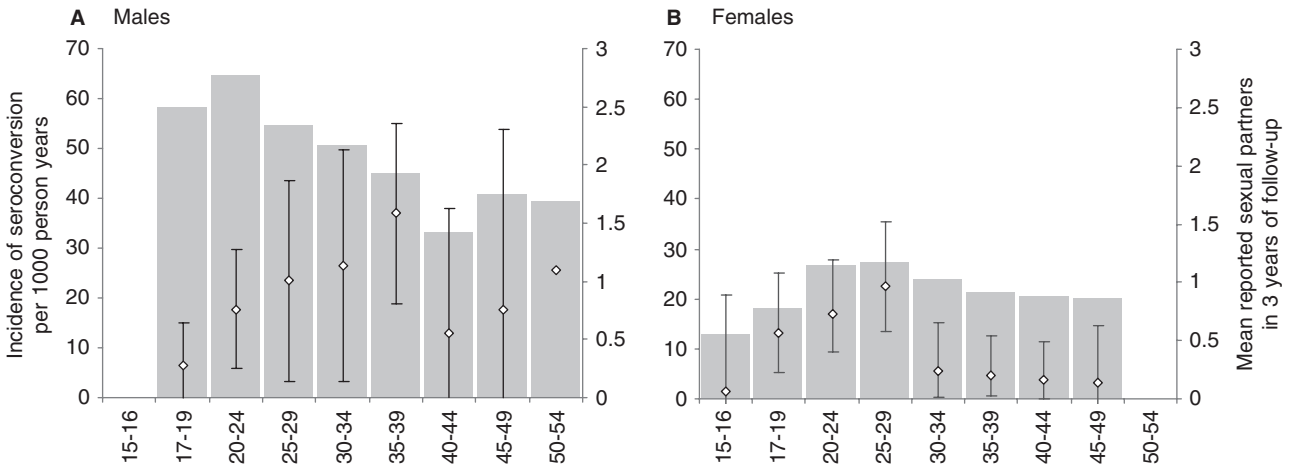


Figure 1 Incidence of HIV seroconversion and mean number of reported sexual partners for males and females, stratified by age. Gray bars are mean sexual partners in three years of follow-up. Circles and bars are point estimates and 95% CIs of HIV incidence by age group

increase in prevalence) were predictors of incident infection. Male circumcision was not associated with HIV incidence ($P = 0.9$) though only 3% of men in this population reported as having been circumcised.

For women, predictors of incident infection were similar to those for men: with multiple partners ($RR = 3.2$), experiencing genital sores ($RR = 2.0$), and having an unwell partner ($RR = 2.0$). In addition, multiple partners in the last month was an independent predictor ($RR = 8.8$), even after controlling for total number of partners in the follow-up period. Dry sex, reportedly practiced by 24% of women, was not associated with HIV incidence ($P = 0.4$).

Between baseline and follow-up the frequency of these determinants declined: multiple partners (men 40% to 21%; women 6% to 3%) and having an STI in the last year (men 13% to 6%, women 35% to 16%) ($P < 0.0001$ for all in chi-squared tests). We did not have a measure of partner's heath at baseline for comparison.

Underlying determinants

In men, HIV incidence increased with onset of sexual activity in young age groups, falling sharply after age group 35–39, then slowly rising again (perhaps because of widows and divorcees acquiring new partners and married men starting second families) although numbers are small (Figure 1). A number of socio-demographic factors were associated with HIV incidence in crude models: being unmarried (protective), being widowed or divorced, living in a subsistence farming area (protective, relative to towns), being a student (protective) and lower socioeconomic status.[6] Increased mobility was associated with incidence in a number of measures: living outside the study site for more then one month, previously living in the countryside and visiting a bar. A number of personal beliefs about condom use acceptability and ways to avoid contracting HIV were associated with incidence. Men who acknowledged that a relative died of AIDS had reduced incidence of infection. Including the proximate determinants (as in Table 2 and above) had little effect aside from slightly reducing the statistical power and parsimony of the models, suggesting that the underlying determinants are measuring different aspects of proximate risk from those measured directly (Table 4).

For women, incidence was observed to rise up the 25–29-year age group, then fall sharply and stay at a low level which parallels the mean number of sex partners (Figure 1). Being widowed or divorced, having secondary school or higher education, being a member of a traditional church and attending a bar were associated with higher incidence, as were beliefs about marriage, beer drinking and getting paid for sex. Only the effect estimate for attending a bar moved substantially towards the null when controlling for the proximate determinants. For women, attending a bar is highly associated with multiple sex partners, which explains the loss of significance of the underlying determinant in the full model (Table 5).

Estimating PAF by approximating R_t

We estimate the reproductive number for men (R_{tm}) was 0.78 and for women (R_{tw}) 0.51 (Figure 2). This suggests a declining epidemic, as has been observed in this cohort.[6] For men, we estimate that having multiple sexual partners accounts for approximately one third of the total reproductive number $[(1-R_{ti})/R_t = 0.35]$, while the other identified proximate determinants (genital sores and having an unwell partner) accounted for only a small proportion of transmission, due to the relative rarity of these factors. For women, no factor had a substantial impact (>10%), again owing to the low reported prevalence of the proximate determinants. Underlying determinants for men

Table 4 Underlying determinants, Males

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Partially adjusted RR[b]	Fully adjusted RR[c]
Age					
15–16					
17–19	5/772	6.4	0.4 (0.1–0.9)	0.2 (0.1–1.1)	0.7 (0.3–2)
20–24	26/1418	18.2	1*	1	1
25–29	23/979	23.2	1.3 (0.7–2.2)	1.1 (0.6–2)	1.1 (0.6–2)
30–34	13/449	28.6	1.6 (0.8–3)	1.3 (0.6–2.6)	1.1 (0.6–2.4)
35–39	16/352	36.6	2 (1.1–3.7)*	1.6 (0.8–3.2)	1.5 (0.8–3)
40–44	5/283	12.9	0.7 (0.3–1.8)	0.5 (0.2–1.4)	0.5 (0.2–1.5)
45–49	5/293	17.5	1 (0.4–2.5)	0.6 (0.2–1.8)	0.6 (0.2–1.7)
50–54	5/195	25.6	1.4 (0.5–3.7)	1.2 (0.4–3.1)	1.2 (0.4–3.4)
Socioeconomic site type					
Town	20/651	30	1	1	1
Estate	42/2091	19.9	0.7 (0.4–1.1)	0.6 (0.3–1.1)	0.5 (0.1–3)
Subsistence farming area	22/1547	14.2	0.5 (0.3–0.9)*	0.7 (0.3–1.3)	0.4 (0–4.3)
Roadside business centre	14/623	22.4	0.7 (0.4–1.5)	1.2 (0.6–2.6)	0.7 (0.1–6.1)
Marital status					
Never married, virgin	0/410	0			
Never married, not virgin	17/1442	11.7	0.5 (0.3–0.8)**	0.5 (0.3–1.1)	0.4 (0.2–0.9)*
Widowed/Divorced	11/215	50.9	2.1 (1.1–4.0)*	1.9 (0.9–4.0)	1.8 (0.8–3.7)
Married	70/2844	24.4	1***	1***	1*
Type of employment					
Unemployed	19/1084	17.4	0.8 (0.5–1.3)	1 (0.5–1.8)	1 (0.6–1.8)
Student	2/563	3.8	0.2 (0–0.7)*	0.7 (0.2–2.6)	0.5 (0.2–1.8)
Professional	0/188	0			
Self-employed	2/197	20.7	0.9 (0.2–3.9)	1.1 (0.3–4.7)	0.9 (0.2–4.1)
Skilled labourer	21/587	35.4	1.6 (1–2.7)	1.6 (0.9–2.7)	1.5 (0.9–2.5)
Manual/unskilled	53/2390	21.9	1**	1*	1*
Secondary or higher education					
No	33/1540	21.4			
Yes	65/3411	19.0			
Church					
Christian	48/22592	18.5			
Traditional	20/809	24.6			
Apostolic[c]	13/535	24.2			
Other/none	17/968	21.0			
Attended bar in last month					
No	33/2495	13.1	1	1	1
Yes	65/2417	26.6	2 (1.3–3.1)**	1.6 (1–2.5)*	1.4 (0.9–2.3)

(continued)

(going to bars, not believing that HIV can be avoided by sticking to one partner and using condoms, and not acknowledging that a family member died of AIDS) may contribute to transmission [$(1-R_{ti})/R_t = 0.29$, 0.28, and 0.14, respectively]. However, since we presume that these factors work through proximate determinants, it is incorrect to think that by hypothetically removing the underlying determinant,

Table 4 Continued

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Partially adjusted RR[b]	Fully adjusted RR[c]
Lived elsewhere					
No	38/2466	15.6	1*		1
City or town	10/557	17.7	1.1 (0.6–2.3)	1.1 (0.5–2.5)	1 (0.5–2.2)
Countryside	50/1888	26.2	1.7 (1.1–2.6)**	1.7 (1–2.9)*	1.7 (1.0–3.0)*
Lived outside study site for 1 month or more in 3 years of follow-up					
No	89/3310	17.5	1	1	1
Yes	39/1599	23.9	1.4 (0.9–2.1)	1.6 (1–2.6)*	1.3 (0.6–3.0)
Know a relative who died of AIDS					
No	68/2938	22.9	1	1	1
Yes	30/1974	14.7	0.6 (0.4–0.9)*	0.6 (0.4–0.9)*	0.5 (0.3–0.9)**
Agreed: One can avoid HIV by sticking to one partner or always using condoms					
No	12/309	38.5	1	1	1
Yes	84/4803	18.3	0.5 (0.3–0.9)*	0.3 (0.2–0.7)**	0.3 (0.2–0.6)**
Agreed: More likely to die from an accident than AIDS					
No	20/649	29.4	1	1	1
Yes	78/4242	18.2	0.6 (0.4–1)*	0.6 (0.3–1)	0.7 (0.4–1.2)
Agreed: Condom use within marriage widely accepted					
No	61/2562	24	1	1	1
Yes	34/2255	14.1	0.6 (0.4–0.9)*	0.6 (0.4–1)*	0.7 (0.4–1.1)
Agreed: Condoms reduce the pleasure of sex					
No	36/2301	14.9	1	1	1
Yes	61/2544	23.9	1.6 (1.1–2.4)*	1.5 (0.9–2.3)	1.6 (1.0–2.5)*
Agreed: Drinking beer is an essential activity for men					
No	30/2216	14.9	1	1	1
Yes	68/3635	23.9	1.6 (1.1–2.4)*	1.8 (1.1–3)*	1.9 (1.1–2.5)
Agreed: I have partner who has other partners but does not always use condoms					
No	85/4557	18.2	1	1	1
Yes	13/3517	37.1	2 (1.1–3.7)*	1.1 (0.7–1.8)	1.6 (0.8 – 3.0)
Agreed: My partner would not use condoms with me on a regular basis					
No	52/3123	16.6	1	1	1
Yes	45/1772	25.3	1.6 (1.1–2.4)*	1.1 (0.7–1.7)	1.0 (0.7–1.7)
Socioeconomic group[d]					
Per quintile			0.86 (0.74–0.98)*	0.89 (0.77–1.05)	0.90 (0.77–1.06)
Agreed: I pay for sex because my friends do and because they encourage me					
No	87/4556	19.1	1		
Yes	11/345	31.9	1.7 (0.9–3.1)		
Circumcised					
No	92/4634	19.8	1		
Yes	6/314	19.1	0.9 (0.4–2.1)		

(continued)

Table 4 Continued

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Partially adjusted RR[b]	Fully adjusted RR[c]
Cared for someone with AIDS					
No	66/3271	20.1	1		
Yes	32/1680	19.1	0.9 (0.6–1.3)		
Agreed: These days most married men are faithful to their wives					
No	46/2336	19.7	1		
Yes	52/2456	20.4	1.0 (0.7–1.5)		

[a]Seroconversions per 1000 person-years.
[b]Adjusted in Poisson regression models for other identified underlying determinants
[c]Adjusted in Poisson regression models for other identified underlying determinants and proximate determinants (Table 2).
[d]Socioeconomic quintile was based on an cumulative index of 11 household-level variables on asset ownership, education and employment of head.
[e]Followers of Apostolic faiths are uncommon in the study areas and do not include the main Apostolic group (Marange) so Apostolics in the study areas may be atypical.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

some proportional amount of transmission would necessarily be averted. The estimated effect on the reproduction number was consistently greater than the conventional PAF but neither measure accounted for a large proportion of the HIV transmission.

Discussion

We identified three variables (number of sexual partners, having an STI and having an unwell partner) that capture the main aspects of the proximate determinants framework: individual behaviour, partnership characteristics and the probability of transmission. Total number of reported sexual partnerships was the individual behavioural measure predictive of incident infection. Suffering an ulcerative STI was the best marker for transmission probability. And, having an unwell partner (and the local prevalence in the opposite sex for men) can be seen as a limited proxy measure for contact with an infected individual, which, of course, is a causal necessity in the sexual transmission of HIV.

Some underlying determinants—if related to sexual behaviour or networks—would be expected to be associated with HIV incidence in univariate analysis. But if the proximate determinants fully captured the mechanism through which the underlying determinants worked, they should no longer carry an effect in fully parameterized models. An example of this is bar attendance, which itself is not a cause of infection. Attending a bar was a marker for infection in women, but after controlling for proximate determinants—i.e. number of partners—bar attendance was no longer a predictor. However, there was surprising residual importance of variables like traditional religion and secondary education (for women) acknowledging that a family member died of AIDS, beliefs about condom use (for men) and beliefs about beer drinking (both sexes). Men who had a prior residence in rural areas or who lived outside their site of residence during the survey period had higher rates of infection. Indeed, these variables may be markers for position in a sexual network, rather than specific risk behaviour. Earlier in the epidemic, mobility was thought to be an important driver of the spread of HIV into rural areas.[17] In Manicaland, migrants themselves did not have higher prevalence at baseline,[17,18] but mobility may continue to play an important role by connecting the sexual network of different geographical sites.

Loss to follow-up was ~40% over the 3-year inter-survey period. Although this figure is similar to other population-based cohorts in Africa and followed-up migrants do not have higher risk of HIV,[17] there is potential for unmeasured differences between those lost to follow-up and those re-interviewed. STD status was based on self-reported symptoms, rather than biological testing, which may have resulted in under-reporting, recall bias or inclusion of non-infectious causes of genital complaints. Participation rates were high, but it remains possible that difficult-to-reach groups, for example female sex workers, were poorly identified in the survey. Therefore the study population may not be completely representative of the population as a whole. Perhaps the most important limitation here is that the serostatus of partners were not known. In the late stage of this epidemic, serostatus of long-term partners may be the most important single individual determinant of risk.

Because of the long incubation of AIDS, prevalent HIV infections are a product of many years of risk, so longitudinal studies on incidence are especially important for understanding the contemporary epidemiology. However, even in this analysis of a prospective cohort, the proximate determinants failed to account for a large proportion of transmission. The proximate determinants framework has proven a very useful concept for studying fertility and child survival.[19,20]

Table 5 Underlying determinants, females

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Partially adjusted RR[b]	Fully adjusted RR[c]
Age					
15–16	2/282	6.4	0.3 (0.1–1.1)	0.4 (0.1–1.7)	0.4 (0.1–1.7)
17–19	18/869	20.7	0.9 (0.5–1.5)	1 (0.5–1.8)	0.9 (0.5–1.7)
20–24	28/1147	24.4	1***	1**	1**
25–29	33/1037	31.6	1.3 (0.8–2.1)	1.3 (0.8–2.2)	1.3 (0.8–2.3)
30–34	9/861	10.4	0.4 (0.2–0.9)	0.4 (0.2–0.8)	0.6 (0.2–1.4)
35–39	10/1114	9.0	0.4 (0.2–0.8)	0.5 (0.2–1)	0.8 (0.3–2.1)
40–44	8/1064	7.5	0.3 (0.1–0.7)	0.5 (0.2–1.1)	0.8 (0.3–2.7)
45–49	5/781	8.1	0.3 (0.1–0.9)	0.4 (0.2–1.2)	0.8 (0.2–3.2)
Socioeconomic site type					
Town	13/603	22.2	1		
Estate	39/1818	21.8	1 (0.5–1.8)		
Subsistence farming area	43/3159	13.8	0.6 (0.3–1.2)		
Roadside business centre	18/1610	11.6	0.5 (0.3–1.1)		
Marital status					
Married	74/5143	14.7	1*	1	1
Never married, virgin	7/658	10.7	0.7 (0.3–1.6)	0.5 (0.2–1.2)	0.6 (0.1–3.3)
Never married, not virgin	5/310	16.1	1.1 (0.4–2.7)	0.7 (0.3–1.8)	0.7 (0.2–1.8)
Widowed/Divorced	27/1079	18.3	1.7 (1.1–2.7)*	1.5 (1.0–2.5)	1.5 (0.8–3.7)
Secondary or higher education					
No	35/3693	9.8	1	1	1
Yes	78/3498	22.4	2.3 (1.5–3.4)***	1.4 (1.1–1.8)**	1.4 (1.1–1.7)*
Church					
Christian	68/5056	13.4	1	1*	1*
Traditional	7/171	40.8	2.9 (1.3–6.3)**	3.4 (1.3–6.5)**	2.9 (1.2–6.7)*
Apostolic[d]	17/904	18.0	1.4 (0.7–2.2)	1.1 (0.6–1.8)	1.0 (0.5–1.6)
Other/none	21/1055	19.9	1.4 (0.9–2.3)	1.3 (0.8–2.2)	1.3 (0.8–2.2)
Attended bar in last month					
No	104/6927	15.3	1	1	1
Yes	9/263	34	2.2 (1.1–4.4)*	2.5 (1.2–5)*	1.5 (0.7–3.2)
Agreed: These days most married men are faithful to their wives					
No	71/3901	18.7	1	1	1
Yes	39/3165	12.6	0.7 (0.5–1)*	0.7 (0.5–1.1)	0.8 (0.5–1.2)
Agreed: Drinking beer is an essential activity for men					
No	57/3901	14.9	1	1	1
Yes	55/2752	23.9	1.5 (1.1–2.2)*	1.6 (1.1–2.4)*	1.5 (1.1–2.3)*
Agreed: I get paid for sex because my friends do and because they encourage me					
No	103/6974	14.7	1	1	1
Yes	10/198	50.3	3.4 (1.7–6.5)***	1.0 (0.4–2.7)	0.6 (0.2–1.7)

(continued)

Table 5 Continued

	Seroconversions/person years at risk	Rate[a]	Unadjusted RR	Partially adjusted RR[b]	Fully adjusted RR[c]
Type of employment					
Unemployed	54/3822	16.4	1		
Student	9/775	11.6	0.7 (0.2–3.2)		
Professional	2/96	20.7			
Self-employed	0/82	0			
Skilled labourer	2/112	17.8			
Manual/unskilled	46/2837	16.2	0.8 (0.2–3.2)		
Lived elsewhere					
No	45/2730	16.4	1		
City or town	15/1090	13.7	0.8 (0.5–1.4)		
Countryside	53/3364	17.8	1.0 (0.6–1.4)		
Lived outside study site for 1 month or more in 3 years of follow-up					
No	73/4710	15.5	1		
Yes	39/2437	16.0	1.0 (0.7–1.5)		
Know a relative who died of AIDS					
No	63/3415	18.4	1		
Yes	50/3770	13.2	0.7 (0.5–1.0)*		
Know a relative who died of AIDS					
No	4/345	11.6			
Yes	109/6830	16.0	1.4 (0.5–3.7)		
Know a relative who died of AIDS					
No	26/1262	20.6	1		
Yes	85/5174	14.9	0.7 (0.5–1.1)		
Know a relative who died of AIDS					
No	64/4098	15.6	1		
Yes	47/2559	18.4	1.2 (0.8–1.7)		
Know a relative who died of AIDS					
No	54/3329	16.2	1		
Yes	49/2981	16.4	1.0 (0.7–1.5)		
Know a relative who died of AIDS					
No	88/5636	15.6	1		
Yes	22/1333	16.4	1.1 (0.7–1.7)		
Know a relative who died of AIDS					
No	57/3323	17.1	1		
Yes	55/3814	14.4	0.8 (0.6–1.2)		

[a]Seroconversions per 1000 person-years.
[b]Adjusted in Poisson regression models for other identified underlying determinants.
[c]Adjusted in Poisson regression models for other identified underlying determinants and proximate determinants (Table 2).
[d]Followers of Apostolic faiths are uncommon in the study areas and do not include the main Apostolic group (Marange) so Apostolics in the study areas may be atypical.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

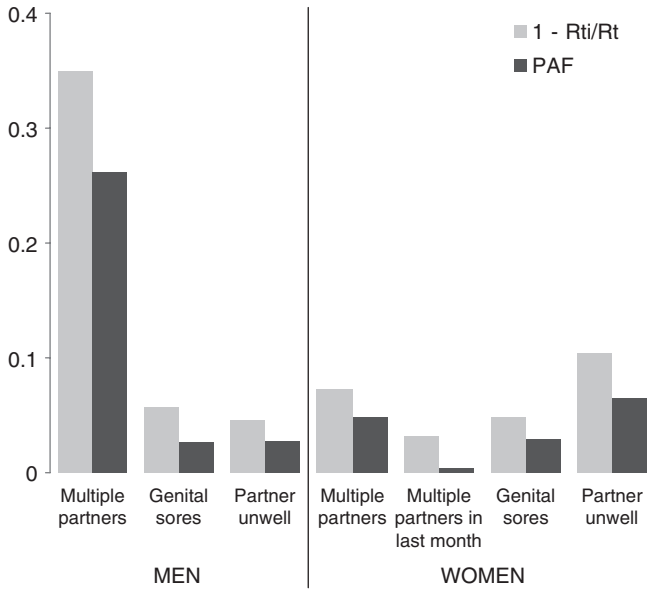


Figure 2 Estimates of population-level impact of proximate determinants. The estimated effect on the reproduction number $[(1-R_{tu})/R_t]$ is greater than the conventional population-attributable fraction (PAF) but neither measures can account for the large proportion of the HIV transmission

But, reconciling the framework to HIV epidemiology may prove difficult because the risk of HIV—as an infectious disease—depends on proximate factors beyond the individual's own behaviour. Removing a risk factor from a population may eliminate proportionally more or less of the disease than would be expected from the risk ratio and the prevalence of the factor depending on how the risk factor affects transmissibility and how individuals with the risk factor mix with the population.[16] Thus, the relative risk of a given factor is dependent on the magnitude and stage of the epidemic. Any empirical study is limited in its follow-up time and, therefore, fails to fully describe sexual transmission as a result of the lifetime sexual network of subjects and their partners.

That our analysis failed to identify the influence of underlying determinants on proximate determinants suggests that the latter are poorly measured. However, this may be because many infections are occurring in those with no discernable risks, as well as infection having spread through the sexual partner networks to those at the ends of chains of transmission who are otherwise typical.

Comparing risk factors for incidence with risk factors for prevalence may elucidate which factors are necessary for an epidemic and which change as the epidemic goes through different phases (Table 6). The main proximate determinants of incidence are similar to those previously found to be risk factors for prevalent infection in this population: number of sexual partners, presence of STI cofactors and (for men) local HIV prevalence in women.[8] However, certain underlying determinants are different. For example, being widowed was a predictor of prevalence, whereas being divorced or separated was a predictor of incidence, suggesting that widow(er)s were infected in marriage, while divorcees were at heightened risk as they acquire new partners. Beliefs, which were not predictors of prevalent infection, were associated with incidence. 'Protective beliefs' included acknowledging that a relative died of AIDS, believing one can avoid AIDS with condoms and fidelity and that condoms did not reduce the pleasure of sex (for male respondents) and beer drinking was not an essential activity for men (for female respondents). All of these protective beliefs increased in frequency between the baseline and follow-up survey. For men, the proximate determinants of multiple partners in last month and genital sores also decreased in frequency between baseline and follow-up. These substantial reductions in risk behaviour were not a result of the intervention trial that was ongoing during the study period,[6] but rather a combination of higher mortality in risky populations and, importantly, individual reductions in risk behaviour including increased condom use, reductions in casual partnerships and delayed first sex.[6]

One counter-intuitive finding for prevalence at baseline (that consistent condom use was a predictor of infection) was not found with incidence. This may suggest that condom use is associated with past risk or that infected people are motivated to use condoms to protect their partner(s). A paradoxical finding was that, amongst women, secondary/higher education was associated with lower risk for prevalence but higher risk for incidence. This contradicts most other studies that show a trend towards lower rates of infection in educated groups as the epidemic progresses.[21,22] This finding requires further investigation.

Risk factor studies for incident infection have not been plentiful, but the published studies have largely identified the same proximate determinants as were found here. Numbers of sexual partners and presence of other STIs (as measured by genital ulcer or urethral discharge) are nearly universal risk factors found in Ugandan,[1,23] Tanzanian[2,24] and Zimbabwean[3] cohorts. Being the victim of forced sex for women,[25,26] and not being circumcised[27,28] for men have been identified as risk factors elsewhere. Perhaps unsurprisingly, distal socio-demographic factors, such as religion are less consistent between studies, apart from marriage which is generally protective.[1,4,23,24]

Our estimates of population impact $[(1-R_{tu})/R_t]$ tended to be greater than those for the PAF, as it estimates the influence on the biological parameter R—the basic reproductive number. This measure approximated the impact of relative susceptibility and, to some extent, accounts for changing dynamics, such as the changing size of the risk groups in follow-up. Methods that conceptualize the population-level impact of a risk factor in terms of the reproductive number R

Table 6 Prevalence and risk of proximate and underlying detriments for HIV incidence in the follow-up period and at baseline survey

	Risk factors for HIV prevalence			Risk factors for HIV incidence				
	Risk period	aOR	Prevalence at baseline	Risk period	aRR	Prevalence at baseline	Prevalence at follow-up	**Risk for**
Males								
Proximate								
Multiple partners	Lifetime	3.6–7.3	84%	3 years	2.8	21%	40%	I + P
Genital sores	Lifetime	3.6	28%	Last year	2.7	5%	13%	I + P
Community HIV prevalence	Current	1.1	NA	Current	1.2	NA	NA	I + P
Multiple partners	Last month	NS		Last month	8.8	4%	9%	I
⩾ 10 years older than partner	Current	2.1	16%	Current	NS			P
Underlying								
Age	Current	Older	NA	Current	Older	NA	NA	I + P
Divorced/separated	Current	2.5	5%	Current	4	4%	4%	I + P
Previously lived in countryside	Lifetime	NS		Lifetime	1.7	38%	43%	I
Know a relative who died of AIDS	Current	NS		Current	0.5	40%	26%	I
Does not believe can avoid HIV by sticking to one partner or always using condoms	Current	NS		Current	2.6	6%	9%	I
Believes condoms reduce the pleasure of sex	Current	NS		Current	1.6	53%	55%	I
Widowed	Current	5.9	1%	Current	NS			P
Skilled, self-employed or professional	Current	0.6	19%	Current	NS			P
Visited bar	Last month	1.3	56%	Last month	NS			P
Females								
Proximate								
Multiple partners	Lifetime	2.2–8.6	36%	Last 3 years	3.2	6%	3%	I + P
Genital sores	Last year	NS		Last year	2	DNC	6%	I
Partner unwell		DNC		Current	2	DNC	13%	I
Multiple partners	Current	0.4	2%	Current	NS			P
Community HIV prevalence	Current	1.6		Current	NS			P
Experienced discharge	Lifetime	1.5	36%	Last year	NS			P
Suspects long-term partner has other partners	Current	1.5	26%	Current	NS			P
Long-term partner aged 25 to 34	Current	2.3	35%	Current	NS			P

(continued)

Table 6 Continued

	Risk factors for HIV prevalence			Risk factors for HIV incidence				
	Risk period	aOR	Prevalence at baseline	Risk period	aRR	Prevalence at baseline	Prevalence at follow-up	**Risk for**
Long-term partner ⩾ 10 years older	Current	1.4	27%	Current	NS			P
Consistent condom use in recent partnership	Last 2 weeks	1.4	7%	Last 2 weeks	NS			P
STI	Lifetime	2.3	2%	Last month	NS			P
Underlying								
Age	Current	20–29 years	36%	Current	20-29 yrs	36%	31%	I + P
Divorced/separated	Current	1.6	14%	Current	2.6	8%	11%	I + P
Secondary/higher education	Current	0.8	47%	Current	1.4	49%	52%	I + P
Believes drinking beer is an essential activity for men	Current	NS		Current	1.5	41%	54%	I
Traditional religion	Current	NS		Current	2.9	2%	3%	I
Town residence	Current	2.1	15%	Current	NS			P
Lived in area 0 to 9 yrs	Current	1.3	44%	Current	NS			P
Had HIV test	Lifetime	1.4	6%	Lifetime	NS			P
Widowed	Current	4	9%	Current	NS			P

NS ($P > 0.05$). NA, Not applicable since variable is not categorical. DNC, Data not collected at baseline survey or not collected in a directly comparable manner. I, Incidence; P Prevalence.
aOR adjusted odds ratio, aRR adjusted incidence rate ratio.

may be more satisfactory than the PAF since they are well grounded in the theory of infectious disease epidemiology.[8] These methods do not assume independence of events, an important violation of directly transmitted diseases of assumptions underlying PAF calculations. However, these *R* methods still do not account for how the risk factor affects infectiousness through concurrency, variable levels of virus shedding or tendencies to mix with susceptible groups. In particular, the population impact of STIs, which increase both susceptibility and infectiousness,[29] will be underestimated using either PAF or *R* methods.

Despite these limitations in the framework, the empirical data and statistical approaches to test the framework, this study enhances the evidence base for understanding the contemporary HIV epidemic in Zimbabwe and for implementing prevention activities. Reported 'high risk' behaviours were quite uncommon, especially in women, highlighting the degree to which the epidemic has become generalised. Indeed, ~70% of infections occurred amongst married men and women.

Having multiple sexual partners remains the key proximate risk factor, although an unwell partner has also emerged as important. The change in underlying risk factors illustrates the barriers to further reductions in HIV incidence, particularly the role of beer drinking in increased risk and the protective effect of variables suggesting an acknowledgement of HIV risk. Most striking is our limited ability to identify factors contributing to the epidemic. This may be because we have been unable to measure some risks, but identifying risk factors assumes that those acquiring infections are somehow different from others who do not acquire infections. That they are not suggests that in this generalized epidemic there is little difference in individual characteristics and behaviours between those who acquire infection and those who do not.

Acknowledgements

We thank the people of Manicaland, Zimbabwe for their time and commitment to the study, all of the field work staff of the Manicaland HIV/STD Prevention Project and the ALPHA network for helpful contributions. We also thank Christophe Fraser for sharing ideas leading to the development of the CRN methods.

Conflict of interest: None declared.

KEY MESSAGES

- The HIV epidemic may have begun to decline in Manicaland, Zimbabwe, but the incidence of new infections remains high at ~2% and 1.5% per year for men and women, respectively. We used an *a priori* framework to identify the best measures of HIV risk of individual behaviour, partnership characteristics and the probability of transmission.
- Multiple sexual partners, having an unwell partner, and reporting another sexually transmitted disease were identified as the best indicators of these proximate determinants. However, these individual risk factors did not explain a large proportion of new infections at the population level, suggesting that at this late stage of the epidemic there is little difference in readily identifiable individual characteristics between those who acquire infection and those who do not.

References

[1] Wawer MJ, Sewankambo NK, Berkley S *et al*. Incidence of HIV-1 infection in a rural region of Uganda. *Br Med J* 1994;**308:**171–73.

[2] Senkoro KP, Boerma JT, Klokke AH *et al*. HIV incidence and HIV-associated mortality in a cohort of factory workers and their spouses in Tanzania, 1991 through 1996. *J Acquir Immune Defic Syndr* 2000;**23:**194–202.

[3] Mbizvo MT, Latif AS, Machekano R *et al*. HIV seroconversion among factory workers in Harare: who is getting newly infected? *Cent Afr J Med* 1997;**43:**135–39.

[4] McFarland W, Mvere D, Katzenstein D. Risk factors for prevalent and incident HIV infection in a cohort of volunteer blood donors in Harare, Zimbabwe: implications for blood safety. *Aids* 1997;**11(Suppl 1):**S97–102.

[5] UNAIDS. *Report on the global AIDS epidemic*. Geneva: Joint United Nations Programme on HIV/AIDS; 2006.

[6] Sherr L, Lopman B, Kakowa M *et al*. Voluntary counselling and testing: uptake, impact on sexual behaviour, and HIV incidence in a rural Zimbabwean cohort. *Aids* 2007;**21:**851–60.

[7] UNAIDS. *Evidence for HIV Decline in Zimbabwe: A Comprehensive Review of the Epidemiological Data*. Geneva, Switzerland; 2005.

[8] Gregson S, Garnett GP, Nyamukapa CA *et al*. HIV decline associated with behavior change in eastern Zimbabwe [see comment]. *Science* 2006;**311:**664–66.

[9] Central Statistical Office and Macro International Inc. Zimbabwe Demographic and Health Survey 2005–2006: Preliminary Report. Harare, Zimbabwe and Calverton, Maryland USA; 2006.

[10] Wasserheit JN, Aral SO. The dynamic topology of sexually transmitted disease epidemics: implications for prevention strategies. *J Infect Dis* 1996;**174(Suppl 2):**S201–13.

[11] Boerma JT, Weir SS. Integrating demographic and epidemiological approaches to research on HIV/AIDS: the proximate-determinants framework. *J Infect Dis* 2005;**191(Suppl 1):**S61–67.

[12] Poundstone KE, Strathdee SA, Celentano DD. The social epidemiology of human immunodeficiency virus/acquired immunodeficiency syndrome. *Epidemiol Rev* 2004;**26:** 22–35.

[13] Ray CS, Mason PR, Smith H, Rogers L, Tobaiwa O, Katzenstein DA. An evaluation of dipstick-dot immunoassay in the detection of antibodies to HIV-1 and 2 in Zimbabwe. *Trop Med Int Health* 1997;**2:**83–88.

[14] Lewis JJ, Donnelly CA, Mare P, Mupambireyi Z, Garnett GP, Gregson S. Evaluating the proximate determinants framework for HIV infection in rural Zimbabwe. *Sex Transm Infect* 2007;**83(Suppl 1):**i61–9.

[15] Grassly NC, Garnett GP, Schwartlander B, Gregson S, Anderson RM. The effectiveness of HIV prevention and the epidemiological context. *Bull World Health Organ* 2001;**79:**1121–32.

[16] Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology* 1995;**6:**142–51.

[17] Mundandi C, Vissers D, Voeten H, Habbema D, Gregson S. No difference in HIV incidence and sexual behaviour between out-migrants and residents in rural Manicaland, Zimbabwe. *Trop Med Int Health* 2006;**11:**705–11.

[18] Coffee MP, Garnett GP, Mlilo M, Voeten HACM, Chandiwana S, Gregson S. Patterns of movement and risk of HIV infection in rural Zimbabwe. *J Infect Dis* 2005;**191:**S159–67.

[19] Bongaarts J. Framework for analyzing proximate determinants of fertility. *Population Dev Rev* 1978;**4:**105–32.

[20] Mosley WH, Chen LC. An analytical framework for the study of child survival in developing countries. 1984. *Bull World Health Organ* 2003;**81:**140–45.

[21] Hargreaves JR, Boler T. Girl power: The impact of girls' education on HIV and sexual behaviour: Action Aid; 2006.

[22] Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. *Trop Med Int Health* 2002;**7:**489–98.

[23] Quigley MA, Morgan D, Malamba SS *et al*. Case-control study of risk factors for incident HIV infection in rural Uganda. *J Acquir Immune Defic Syndr* 2000;**23:**418–25.

[24] Borgdorff MW, Barongo LR, Klokke AH *et al*. HIV-1 incidence and HIV-1 associated mortality in a cohort of urban factory workers in Tanzania. *Genitourin Med* 1995;**71:**212–25.

[25] Mekonnen Y, Sanders E, Messele T *et al*. Prevalence and incidence of, and risk factors for, HIV-1 infection among factory workers in Ethiopia, 1997–2001. *J Health Popul Nutr* 2005;**23:**358–68.

[26] Kiwanuka N, Gray RH, Serwadda D *et al*. The incidence of HIV-1 associated with injections and transfusions in a prospective cohort, Rakai, Uganda. *Aids* 2004;**18:**342–44.

[27] Gray RH, Kiwanuka N, Quinn TC *et al*. Male circumcision and HIV acquisition and transmission: cohort studies

in Rakai, Uganda. Rakai Project Team. *Aids* 2000;**14:**2371–81.

[28] Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. *PLoS Med* 2005;**2:**e298.

[29] Rottingen JA, Cameron DW, Garnett GP. A systematic review of the epidemiologic interactions between classic sexually transmitted diseases and HIV - How much really is known? *Sexually Transmitted Dis* 2001;**28:**579–97.

Appendix

Estimation of the basic reproductive number in unexposed individuals

Following on from the incidence prevalence ratio described in formula (1), we can calculate the expected number of cases $E[G(t,t+n)_u]$ that would have occurred if the whole population were unexposed to the risk factor. We estimate this by applying the rate of infection in unexposed individuals to the whole population:

$$G(t,t+n) = (1-p)G(t,t+n)_u + p\,G(t,t+n)_u aRR$$

where p is the proportion exposed to the risk factor; $G(t,t+n)_u$ are the incident cases in the unexposed population and *aRR* is the adjusted *RR* (or hazard ratio) from regression analysis, which can be re-arranged to solve for $G(t,t+n)_u$

$$E\big[G(t,t+n)_u\big] = \frac{O[G(t,t+n)]}{[(1-p)+p\times aRR)]}$$

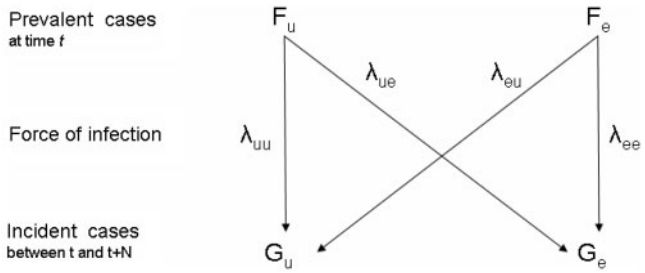
We then estimate the reproductive number for a hypothetical cohort where the risk factor was absent (R_{tu}). R_t was recalculated for those individuals unexposed to the risk factor R_{tu}.

$$Rtu \approx \frac{D}{F(t)_u}$$

In this case, $F(t)_u$ is the number of unexposed cases at time t. Similarly $G(t,t+n)_u$ refers to incident cases, but their exposure status is based on the follow-up period, so a person exposed at baseline can become unexposed and, if infected, contribute to $G(t,t+n)_u$. The ratio of R in the unexposed to R in the whole population is then expressed as in formula (2).

Why the CRN is a conservative estimate

In the proposed estimation of a risk factor's CRN, R_{tu} is calculated as the ratio of incident to prevalent cases that are unexposed to the risk factor. This will always underestimate the CRN because unexposed cases will have been infected by both exposed and unexposed prevalent cases, though we assume they all come from unexposed in the formula for R_{tu}. Diagrammatically:



So, in reality, exposed and unexposed incident cases are generated as follows:

$$G_u = n(\lambda_{uu}F_u + \lambda_{eu}F_e)$$
$$Ge = n(\lambda_{ue}F_u + \lambda_{ee}F_e)$$

However, in an epidemiological study without contact tracing, we cannot observe incidence stratified by exposure status of the infecting person: λ_{uu} and λ_{eu}. We can only observe incidence in the exposed λ_e and in the unexposed λ_e. This requires us to make assumptions about the source of infection. The conservative assumption is that mixing is completely assortative: exposed only infect exposed and unexposed only infect unexposed. We therefore assume that:

$$R_u = R_{uu}$$

which is an underestimation of the true situation where:

$$R_u = R_{uu} + R_{eu}$$

Therefore R_u is smaller than we have approximated. And in turn, the CRN of a risk factor is larger than we have approximated since:

$$\text{CRN} \approx \frac{1-R_{tu}}{R_t}$$

Just as the contribution of risk factors to the PAF can sum to greater than one, the contribution of risk factors to the reproductive number in this formulation can add to more than one, because individuals can have more than one risk and risks can combine to exacerbate the spread of infection.