Tropical Medicine and International Health doi:10.1111/j.1365-3156.2007.01986.x

VOLUME 13 NO 2 PP 162–170 FEBRUARY 2008

The role of context: neighbourhood characteristics strongly influence HIV risk in young women in Ndola, Zambia

Sabine Gabrysch, Tansy Edwards and Judith R. Glynn for the Study Group on heterogeneity of HIV epidemics in African cities

London School of Hygiene and Tropical Medicine, London, UK

Summary

OBJECTIVES To examine the effect of neighbourhood socioeconomic factors on human immunodeficiency virus (HIV) prevalence in young women (aged 15–24 years) in Zambia.

METHODS Re-analysis of a cross-sectional, population-based sero-survey of nearly 2000 adults conducted in 1997/1998 in Ndola, Zambia. Neighbourhood-level socioeconomic status (SES) was defined using the availability of running water and electricity in addition to educational, employment and occupational characteristics of adults older than 24 years. Neighbourhood-level and individual-level risk factors were analysed with a multivariate multilevel logistic regression model using a hierarchical conceptual framework.

RESULTS Young women living in neighbourhoods of lower or middle SES had higher HIV prevalences than those from higher SES neighbourhoods [lower SES: adjusted odds ratio (OR) 2.4, 95% confidence interval (CI) 1.3–4.5, middle SES: adjusted OR 2.4, 95% CI 1.3–4.7]. Young women living near a market were at increased risk of HIV infection (OR 2.9, 95% CI 1.4–5.9), while proximity to a health centre seemed protective (OR 0.4, 95% CI 0.2–1.0). When controlling for neighbourhood factors, better education was a risk factor for HIV infection (OR 1.5, 95% CI 1.0–2.1), although it was not significant in individual-level analysis.

CONCLUSIONS Community-level factors are as important as individual-level factors in determining HIV infection in young women. Confining analyses to individual-level factors ignores the underlying causes and the modifying effect of context on individual behaviour and may even lead to different conclusions concerning the role of individual-level factors.

keywords Human immunodeficiency virus, Africa, women, risk factors, socioeconomic factors, effect modifiers

Introduction

After several decades of epidemiological research focussing on individual-level behavioural and biological risk factors, there has been a resurgence of interest in population-level risk factors, acknowledging the fact that an individual's risk is not only determined by their own behaviour but to a large extent by the context in which they live (Pearce 1996; Susser & Susser 1996).

This population approach builds on Geoffrey Rose's concept of 'sick populations and sick individuals' (Rose 1992). It is now again increasingly appreciated that there are underlying 'causes of causes' which affect individuals directly or constrain the choices they make, and that group-level factors can contain information not, or not easily, captured by individual-level data (Diez-Roux 1998). Acknowledging the role of context can, for instance, help explain the uneven spread of infections through populations and subgroups and why identical individual behaviour can have different outcomes (Aral *et al.* 2005).

In order to take the hierarchical relationships between different factors into account when controlling for confounding variables and for variables on the causal pathway, a conceptual framework is required, as suggested by Victora *et al.* (1997). Such a hierarchical 'proximate-determinants' framework was originally used in the fertility literature and was recently applied to human immunodeficiency virus (HIV) infection by Boerma and Weir (2005). The proximate determinants (e.g. sexual behaviour variables) are the link between biological factors and underlying socioeconomic, sociocultural and demographic determinants.

Few studies on socioeconomic status (SES) and HIV infection in women have looked at population-level factors, compared with the many studies of individual-level factors. Wojcicki (2005) recommended in her review that 'future studies should examine SES at the individual and ecological level' to fill this gap. A study conducted in South Africa found a non-linear relationship between individual and community SES and HIV infection (Johnson & Budlender 2002). Risk was highest there among medium SES individuals in medium SES communities, which had a high proportion of migrant workers. A study in rural Tanzania also found strong community-level effects on HIV risk for variables associated with population mobility and mixing (Bloom *et al.* 2002). Another study in Zimbabwe found that community type was a determinant for HIV infection in women, but not in men (Lewis *et al.* 2006).

Data from the 'Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa' (Buve *et al.* 2001a) have been used to examine the effect of (individual-level) education and socioeconomic factors on HIV prevalence (Hargreaves 2002; Glynn *et al.* 2004). In this analysis, we examine the role of neighbourhood and individual factors on HIV infection in young women aged 15–24 years in Ndola, Zambia. Young women have a very high HIV prevalence and are a key target group for interventions (Laga *et al.* 2001). As there is relatively little mortality in this group, risk factors should be those for acquiring infection rather than those associated with survival.

Methods

Data collection

A population-based sero-survey was conducted in Ndola in 1997/1998 as part of a study on factors which could explain the different prevalences of HIV in East and West African cities. Full details of the methods have been published (Buve *et al.* 2001a). In summary, two-stage sampling was used to obtain a random sample of around 1000 men and 1000 women. Trained interviewers visited households and sought consent to interview all adults aged 15–49 years who slept there the previous night. Interviewers also recorded features of the cluster, such as availability of electricity and running water. Participants were then asked to give blood to test for HIV, herpes simplex virus type 2 (HSV2) and syphilis in addition to urine to test for gonorrhoea and chlamydia.

Laboratory tests

For HIV testing, an enzyme-linked immunosorbent essay (ELISA) was used with confirmation of positive samples with a rapid test. Samples giving discrepant results were tested by Western blot or two further tests. Syphilis was tested using rapid plasma reagin (RPR) and positive samples were tested by *Treponema pallidum* particle-agglutination (TPPA). Individuals positive on both tests were considered to have positive syphilis serology. HSV2 tests were performed with an HSV2 type-specific IgG ELISA (Buve *et al.* 2001a).

Derived neighbourhood-level variables

The 32 clusters sampled in Ndola belonged to 16 neighbourhoods. Five socioeconomic variables were derived to represent various aspects of SES: education, occupation and employment levels among all adults over 24 years of age were each categorized into tertiles. Electricity and water availability was high and therefore categorized into only two groups: uniformly available in the neighbourhood or not.

Using these five derived variables, a composite SES variable was constructed: the lowest category of each variable scored zero credits, the highest scored two and any middle categories scored one. This resulted in a SES composite variable with a score of 0–10, which was grouped into three similarly sized groups: 0–3 = lower, 4–6 = middle, 7–10 = higher SES.

To address concerns that education might be different from other SES variables, neighbourhood classification was repeated excluding the education level of adults. This made very little difference to the groups: only one neighbourhood containing 14 girls would be classified differently (medium instead of low SES). The classification including education level was used in the subsequent analyses.

Another two derived neighbourhood-level variables were created: circumcision level in men (binary, as levels were very low) and HIV prevalence in men. HIV prevalence and education level were age-standardized, as they were strongly associated with age.

Conceptual framework

The conceptual multilevel framework underlying this analysis (Figure 1) is based on Boerma and Weir's proximate determinants framework (2005) for HIV infection. Underlying determinants such as socioeconomic and environmental factors influence the so-called proximate determinants which include sexual behaviour, circumcision and sexually transmitted infections (STI), which in turn have a direct effect on biological mechanisms.

Statistical analysis

Participants with missing data on HIV or on any of the exposures were analysed for systematic differences in other

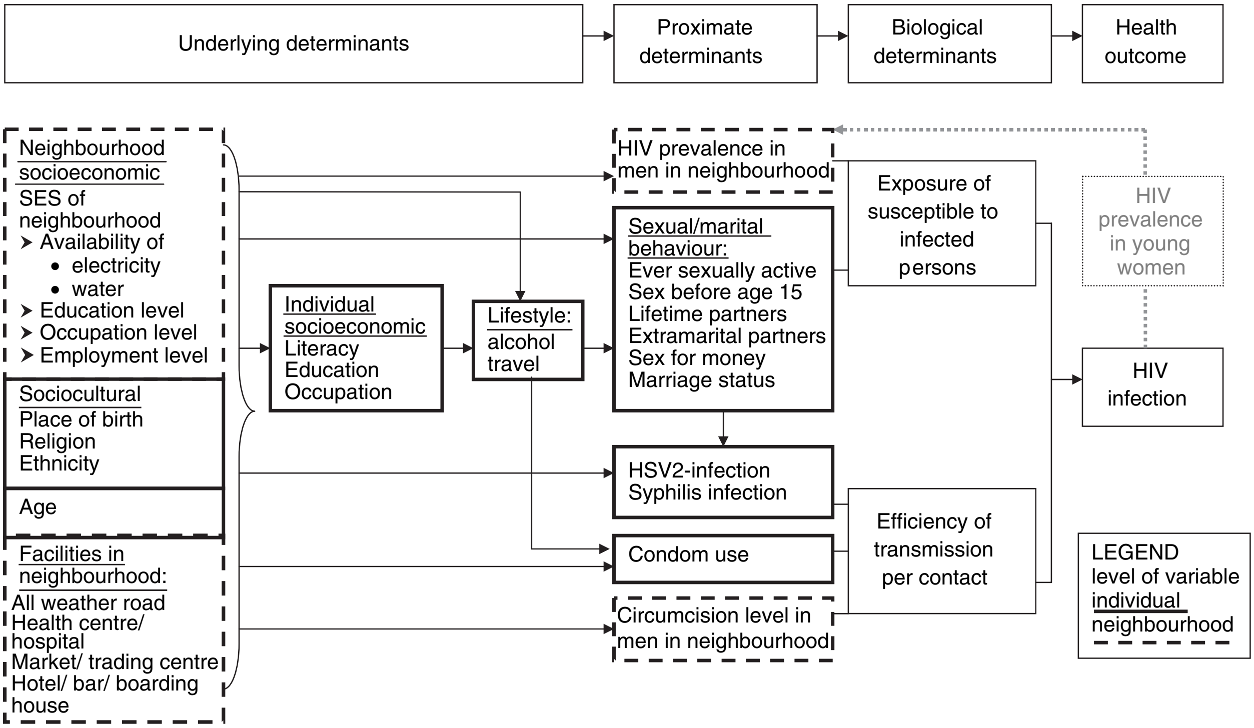

Figure 1 Proximate determinants framework, adapted from Boerma and Weir (2005) for young women in the multicentre study.

variables. All further analysis was then restricted to individuals with known HIV status.

As two-stage sampling was used, the data are potentially clustered on two levels, cluster level and household level. There was no evidence [likelihood ratio test (LRT) $P = 1.0$] of any within-household correlation [intraclass correlation coefficient (ICC) = 9.3×10^{-8}] when looking at HIV infection in young women, but there was some evidence (LRT $P = 0.04$) of within-cluster correlation (ICC = 0.05). Therefore, regression modelling took account of the clustering at cluster level by using two-level random effects models.

The distribution of potential risk factors was described and their association with HIV examined. Age was adjusted for as a linear effect. The distribution of risk factors was compared over the three levels of SES.

Multivariate model building was guided by the hierarchical framework and built stepwise from distal to proximal (Victora *et al.* 1997). Models were built for all women and for sexually active women separately. Separate models were built first for underlying factors only, then including sexual behaviour variables and additionally including STI, as factors that may be on the causal pathway. Controlling for factors that are indeed on the causal pathway should reduce the associations found, to the extent to which they explain these associations. Variables were kept in the model if they reached significance on the 5% level, or on the 10% level if they influenced the odds ratio (OR) of other variables substantially (±15%). When adding proximate stages to the model, those distal variables that had become insignificant were kept in the model in order to show this.

Effect modification by the *a priori* defined factors – age, marital status, SES and male HIV prevalence, was examined using Mantel–Haenszel stratification and tested in the multivariate models with LRT.

Results

Ten per cent of eligible young women (15–24 years) and 25% of eligible men were not interviewed. This was mainly due to not finding them at home despite repeat visits. Few of those who were found refused to be interviewed (Buve *et al.* 2001a).

For 90% (447/496) of the young women who were interviewed, data on HIV status were available. Sixteen per cent of the young women from lower SES neighbourhoods refused testing compared with only 5% in middle SES neighbourhoods and 8% in higher SES neighbourhoods (χ^2 test $P = 0.004$). Young women who agreed to testing were more likely to have completed primary school than those who refused ($P = 0.02$).

The overall HIV prevalence among women aged 15–24 years in Ndola was 28.4% (127/447), with a rise in HIV prevalence from 4.8% in 15-year-olds to 47.4% in 24-year-olds. Neighbourhood SES, having a market nearby, occupation and alcohol use were all associated with HIV (Table 1). Young women from middle and lower SES neighbourhoods had 2.4 (95%CI 1.4–4.3) and 2.3 (95%CI 1.3–4.2) times the odds of HIV infection compared with those from higher SES neighbourhoods. As HIV risk in the non-student occupational groups (housework, sales/service, unemployed) was similar, these were combined into one group.

Several proximate factors were associated with HIV in this analysis (Table 2). Sexual activity, number of lifetime partners and having had a non-marital partner in the previous 12 months were associated with higher HIV prevalence. When adjusting for non-marital partners, the effect of lifetime number of partners was lost and the effect of marital status became significant. Infection with HSV2 or syphilis were also risk factors for HIV infection. Male HIV prevalence in the neighbourhood was associated with HIV prevalence in young women in univariate analysis, but this was lost after adjustment for behavioural factors (OR 1.29, 95% CI 0.8–2.1, $P = 0.31$).

In order to assess which of the individual risk factors are likely to be on the causal pathway between SES and HIV infection in young women, we examined the association of SES with more proximate risk factors (Table 3). Women from lower SES neighbourhoods had lower education levels and they started sexual activity earlier and got married earlier than those from higher SES neighbourhoods. Among those who were sexually active, there was no difference in the number of lifetime partners and non-marital partners in the previous year between the SES strata. The poorer the neighbourhood a young woman came from, the more likely she was to have received money for sex if she had any non-marital partners.

As the neighbourhood SES variable was constructed using education, occupation and employment levels among adults over 24 years of age in the neighbourhood, in addition to electricity and water availability, there is no automatic correlation with the characteristics of the young women under study.

In the multivariate model of underlying risk factors in all young women (Table 4), associations with having a health centre in the cluster (OR 0.41) and higher educational level (OR 1.45) were seen that were not apparent in the univariate analysis. There was some evidence of interaction

Table 1 Underlying risk factors: distribution and age-adjusted effect on HIV prevalence in young women in Ndola

	n	HIV (%)	aOR (95% CI)	LRT *P* value
Cluster-level factors				
Water				
Yes	417	28.5	1	0.90
No	30	26.7	0.94 (0.4–2.4)	
Electricity				
Yes	408	29.7	1	0.10
No	39	15.4	0.44 (0.2–1.2)	
Health centre				
No	385	28.3	1	0.90
Yes	62	29.0	0.96 (0.5–1.9)	
Hotel				
No	250	28.0	1	0.60
Yes	197	28.9	1.14 (0.7–1.9)	
Market				
No	351	25.6	1	**0.02**
Yes	96	38.5	**1.87** (1.1–3.2)	
Neighbourhood-level factor				
SES (composite)				
Higher	140	17.1	1	**0.007**
Medium	156	35.3	**2.41** (1.4–4.3)	
Lower	151	31.8	**2.32** (1.3–4.2)	
Individual-level factors				
Literacy				
Yes	375	29.3	1	0.32
No	71	23.9	0.71 (0.4–1.4)	
Education				
<Primary	158	25.3	1	0.54
Primary	222	30.2	1.30 (0.8–2.2)	
>Primary	67	29.9	1.02 (0.5–2.1)	
Occupation				
Not student	355	33.0	1	**0.04**
Current student	92	10.9	**0.47** (0.2–1.0)	
Ethnic group				
Bemba	291	29.9	1	0.34
Nyanja	85	29.4	0.95 (0.5–1.7)	
Others	71	21.1	0.62 (0.3–1.2)	
Religion				
Protestant	287	26.8	1	0.38
Catholic	124	33.9	1.38 (0.8–2.3)	
Others	36	22.2	0.88 (0.4–2.1)	
Place of birth				
Ndola	248	27.4	1	0.73
Other city	124	32.3	1.13 (0.7–1.9)	
Rural	73	26.0	0.85 (0.5–1.6)	
Alcohol in previous month				
No	389	26.0	1	**0.02**
Yes	58	44.8	**2.09** (1.1–3.8)	
Travel in previous year				
No	228	24.6	1	0.14
Yes	206	33.5	1.39 (0.9–2.2)	

aOR, age-adjusted odds ratio with age as linear effect. aOR shown in bold are significant at the 5% level.
CI, confidence interval; LRT, likelihood ratio test; SES, socioeconomic status.

Table 2 Proximate risk factors: distribution and age-adjusted effect on HIV prevalence in young women in Ndola

	n	HIV (%)	aOR (95% CI)	LRT *P* value
Ever sex				
No	106	7.6	1	**0.002**
Yes	341	34.9	**3.59** (1.6–8.1)	
Ever sexually active women only				
Sex before age 15				
No	272	35.3	1	0.96
Yes	67	32.8	1.02 (0.6–1.8)	
Lifetime partners				
1 partner	164	28.7	1	**0.04***
2 partners	87	37.9	**1.49** (0.8–2.6)	
3 + partners	87	42.5	**1.77** (1.0–3.1)	
Ever married				
No	126	24.6	1	0.12
Yes	215	40.9	1.52 (0.9–2.6)	
Non-marital partner in previous 12 months				
No	274	33.9	1	**0.04**
Yes	67	38.8	**1.86** (1.0–3.4)	
HSV2-infection				
No	167	15.6	1	**<0.001**
Yes	164	54.9	**5.76** (3.4–9.8)	
Syphilis infection				
No	282	30.9	1	**<0.001**
Yes	51	56.9	**3.26** (1.7–6.2)	
Women with non-marital partner only				
Money for sex from non-marital partner				
No	37	43.2	1	0.99
Yes	30	33.3	1.00 (0.3–3.2)	
Condom use with non-marital partner				
No	38	42.1	1	0.11
Yes	27	33.3	0.37 (0.1–1.2)	
Neighbourhood characteristics				
Male HIV prevalence (age-standardized)				
<20%	160	20.6	1	**0.04***
20–27%	171	31.6	**1.67** (1.0–2.9)	
27 + %	116	34.5	**1.95** (1.1–3.5)	
Male circumcision level				
<8%	215	31.6	1	0.15
>8%	232	25.4	0.70 (0.4–1.1)	

aOR, age-adjusted odds ratio with age as linear effect. aOR shown in bold are significant at the 5% level.
CI, confidence interval; LRT, likelihood ratio test.
*LRT for trend.

by age group (LRT $P = 0.05$) on the effect of education, with the elevated HIV risk among the higher educated present only in the younger girls (15–19 years). After additionally adjusting for sexual activity, the effects of SES, being a student and alcohol use on HIV infection were weakened (Table 4).

The multivariate analysis of risk factors among sexually active young women was restricted to those 96% of women with data on all important variables. The same underlying risk factors were identified as in young women on the whole (Table 4). However, SES, alcohol use and currently being a student were slightly less important as some of their action seemed to have been mediated by influencing initiation of sexual activity. Including marital status in the model reduced the effect of SES on HIV infection, while including 'non-marital partner in the previous year' reduced the effect of alcohol use. Additionally adding STI to the model diminished the effect of SES on HIV infection further. Moreover, the protective effect of being a student was reduced by adding HSV2. The effects of market availability and education however got stronger (Table 4). There were no statistically significant interactions between SES and other risk factors.

Discussion

The SES of the neighbourhood was found to be an important determinant of HIV infection in young women in Ndola. Girls from lower or middle SES neighbourhoods had much higher HIV prevalence than girls from higher SES neighbourhoods. Living in a neighbourhood near a market also increased the risk of HIV infection for young women substantially, whereas the presence of a health centre decreased it. The magnitude of these effects was comparable to that of individual-level factors.

Controlling for neighbourhood factors revealed an association between higher individual educational level and HIV infection that had not been apparent before as it had been negatively confounded by neighbourhood SES and current student status. This shows that ignoring population-level factors might lead to different conclusions concerning the role of individual-level factors.

We identified some differences between the women in the different SES neighbourhoods that could help to explain the HIV patterns found. Young women from middle and lower SES neighbourhoods were more likely to be sexually active, married and infected with HSV2 or syphilis than women from higher SES neighbourhoods; these factors were all risk factors for HIV. Their inclusion in the multilevel model reduced the effect of SES by around 20%, suggesting that they may indeed be on the causal pathway. However, even after controlling for proximate factors, girls from poor and middle SES neighbourhoods still remained at nearly twice the odds of HIV infection in comparison with girls from higher SES neighbourhoods. A large proportion of the association of SES and HIV infection thus remains unexplained by the proximate factors measured.

Current students had a lower HIV prevalence than non-students, which seemed to be mediated by students being less likely to drink alcohol, to be sexually active or

Table 3 Association of SES with other risk factors in young women in Ndola, and association of SES with risk factors in men in the neighbourhood

	SES of the neighbourhood (composite variable)			
	Higher	Middle	Lower	*P* value
All young women (*n*):	140	156	151	
Mean age (in years)	19.1	19.9	19.4	0.08[k]
Ever sexually active	**61%**	**86%**	**80%**	**0.009**
Alcohol use last month	13%	16%	10%	0.36
Travel previous year	**55%**	**55%**	**33%**	**<0.001***
Illiterate	**7%**	**13%**	**27%**	**<0.001***
Secondary education	**28%**	**11%**	**7%**	**<0.001***
Current student	28%	21%	14%	0.02*
Ever sexually active (*n*):	86	134	121	
Sex before age 15	**11%**	**22%**	**24%**	**0.03***
Mean age at 1st sex (years)[g]	16.9	15.9	15.6	<0.001[k]
Mean lifetime partners[g]	1.7	1.7	1.6	0.90[k]
Any non-marital partners in last year	23%	16%	21%	0.37
Ever married	41%	68%	74%	<0.001*
HSV2-positive	44%	48%	55%	0.08*
Syphilis positive	10%	15%	19%	0.10*†
Syphilis pos., age 15–19 years	3% (1/29)	11% (6/56)	29% (14/48)	<0.001*
Syphilis pos., age 20–24 years	13% (7/52)	18% (14/77)	13% (9/71)	0.57
Non-marital partner in previous year (*n*):	20	21	26	
Money for sex	15%	43%	69%	<0.001*
Ever condom use with non-marital partner	42%	62%	24%	0.13
Among men in the neighbourhood (n):	169	253	202	
Mean HIV prevalence	20%	26%	23%	0.36
Percentage circumcised	10%	7%	10%	0.34

<TN>Percentages and means shown in bold are significant at the 5% level.
[k]Kruskal–Wallis test, not adjusted for age.
[g]Geometric mean.
*Likelihood ratio test for trend.
†Evidence for interaction by age group (*P* = 0.02).

infected with HSV2, as adding those factors diminished the effect of student status.

Surprisingly, the effects of market and health centre availability in addition to individual education became stronger rather than weaker when adjusting for individual behaviour, STI or male HIV prevalence. It seems that the proximate factors are acting as negative confounders rather than being on the causal pathway.

None of the studies compared in the review by Wojcicki (2005) analysed the interactions between SES and other risk factors. In this analysis, SES was not found to be a significant effect modifier for other risk factors, but power might have been too low to detect any. It seems quite possible that depending on the SES of the neighbourhood, a different set of risk factors might play a role in determining HIV prevalence among young women.

There are several limitations to this study. As in any cross-sectional study, the time sequence of exposures and outcome cannot be easily established. While being a student may be protective, pregnancy leads to school exclusion and drop-out. Similarly, current neighbourhood SES might not have been the aetiologically relevant exposure. Since the study was limited to young women, HIV infection is likely to be recent, but movement between neighbourhoods cannot be excluded.

The 10% of young women who refused HIV testing were more likely than those tested to live in lower SES neighbourhoods and not to have completed primary school. The response rate was very high among young women and lower among men, which could have led to biased estimates of background HIV prevalence among men in the neighbourhood. This and the small numbers of individuals in each neighbourhood limit the ability to detect associations. It is also probable that women have partners from neighbourhoods other than their own, which makes it difficult to measure HIV prevalence in potential partners accurately. It seems likely that this factor accounts for some of the SES effect unexplained by sexual behaviour

Table 4 Multivariate multilevel regression analysis of underlying and proximate risk factors for HIV infection in young women in Ndola

	Adjusted odds ratios (95% CI) (for all other variables in the model)				
	All women (*n* = 447)		Sexually active women (*n* = 331)		
	Underlying factors only	Including whether sexually active	Underlying factors only	Including sexual behaviour	Including sexual behaviour and other STI
Age (linear, per year)	**1.25** (1.1–1.4) $P < 0.001$	**1.21** (1.1–1.3) $P < 0.001$	**1.20** (1.1–1.3) $P = 0.001$	**1.20** (1.1–1.4) $P = 0.002$	**1.15** (1.0–1.3) $P = 0.04$
SES: medium *vs.* higher	**2.44** (1.3–4.5) $P = 0.004$	**2.15** (1.2–4.0) $P = 0.02$	**2.17** (1.1–4.3) $P = 0.03$	**1.97** (1.0–4.0) $P = 0.06$	**1.95** (0.9–4.2) $P = 0.09$
SES: lower *vs.* higher	**2.43** (1.3–4.7) $P = 0.008$	**2.19** (1.1–4.2) $P = 0.02$	**2.49** (1.2–5.2) $P = 0.01$	**2.03** (0.9–4.4) $P = 0.07$	**1.90** (0.8–4.3) $P = 0.13$
Market nearby	**2.85** (1.4–5.9) $P = 0.005$	**2.98** (1.4–6.2) $P = 0.004$	**3.55** (1.6–7.8) $P = 0.002$	**3.62** (1.6–8.1) $P = 0.002$	**4.47** (1.8–11) $P = 0.001$
Health centre nearby	**0.41** (0.2–1.0) $P = 0.05$	**0.42** (0.2–1.0) $P = 0.06$	**0.32** (0.1–0.9) $P = 0.03$	**0.31** (0.1–0.8) $P = 0.02$	**0.26** (0.1–0.8) $P = 0.02$
Current student	**0.47** (0.2–1.0) $P = 0.06$	**0.59** (0.3–1.4) $P = 0.22$	**0.47** (0.2–1.3) $P = 0.15$	**0.43** (0.1–1.3) $P = 0.13$	**0.69** (0.2–2.3) $P = 0.55$
Education (3 groups, linear)	**1.45** (1.0–2.1) $P = 0.05$	**1.47** (1.0–2.1) $P = 0.05$	**1.75** (1.2–2.3) $P = 0.007$	**1.83** (1.2–2.8) $P = 0.005$	**2.15** (1.3–3.5) $P = 0.002$
Alcohol use	**2.21** (1.2–4.1) $P = 0.01$	**2.03** (1.1–3.8) $P = 0.02$	**2.01** (1.1–3.8) $P = 0.03$	**1.82** (1.0–3.5) $P = 0.07$	**1.82** (0.9–3.7) $P = 0.10$
Ever sex	–	**2.67** (1.1–6.2) $P = 0.02$	–	–	–
Ever married	–	–	–	**2.11** (1.0–4.3) $P = 0.04$	**1.73** (0.8–3.8) $P = 0.17$
Non-marital partner last year	–	–	–	**3.06** (1.4–6.7) $P = 0.005$	**2.62** (1.1–6.1) $P = 0.03$
HSV2 positive	–	–	–	–	**5.58** (3.1–10) $P < 0.001$
Syphilis positive	–	–	–	–	**2.16** (1.0–4.4) $P = 0.04$

CI, confidence interval; STI, sexually transmitted infections; SES, socioeconomic status.

variables in the women. Therefore, the role of background HIV prevalence requires further, more detailed study.

Social desirability of answers is likely to have caused misclassification in the sexual behaviour variables. It is known that women are likely to underreport sexual experiences, while men may over-report them (Buve *et al.* 2001b; Glynn *et al.* 2001). Of 106 girls in Ndola who claimed to be virgins, 16 had an STI. Assuming that these 'virgins' had the same STI prevalence (46%) as declared sexually active girls with the same age distribution would imply that a third of all declared virgins were not really virgins. Under-reporting of sexual behaviour will lead to underestimation of the role these factors play on the causal pathway of more distal factors. The small number of women reporting non-marital partners also hampered the analysis of specific risk factors in this subset.

Information on sociodemographic and socioeconomic factors should be more reliable than on behavioural factors. However, SES is a construct out of various aspects which also change over a lifetime; hence, any definition will always have its limitations (Braveman *et al.* 2005). As socioeconomic factors were not the main focus of the original study, data on these are limited. It is therefore possible that some important aspects were missed.

Similar associations of low SES with HIV have been described for young women in Kenya (Zutavern 2001; Hargreaves 2002) and elsewhere (Wojcicki 2005) but mostly on the individual level and without identifying any factors on the causal pathway. It has also been described that young women living in trading villages are at higher HIV risk than those in rural villages (Konde-Lule *et al.* 1997) and that HIV prevalence is higher in economically active communities (Johnson & Budlender 2002). Our analysis indicates that market proximity might also be a risk factor inside towns. In rural Tanzania (Bloom *et al.* 2002), strong associations of HIV infection with several community-level factors related to trading, mobility and mixing were weakened in women when including individual-level factors.

It has been suggested that early in the HIV epidemic, educated people of higher SES, who could travel, were predominantly affected and that this trend may reverse as the epidemic becomes more wide-spread and prevention measures are starting to have an effect (Over & Piot 1996; Gregson *et al.* 2001; Hargreaves & Glynn 2002). Several studies during the 1990s seemed to confirm this (Kilian *et al.* 1999; Crampin *et al.* 2003; De Walque *et al.* 2005). In Ndola, the turning point might have been around the time of the multicentre study 1997/1998. A previous analysis of the Ndola data (Glynn *et al.* 2004) did not find any association between education and HIV prevalence, but did not account for neighbourhood-level SES. More recent data from Zambia (Fylkesnes *et al.* 2001; Michelo *et al.* 2006) showed a decline in HIV prevalence among higher educated young people, while prevalence in less educated youths was stable or rising.

Our results show that individual-level and population-level factors are quite distinct concepts and can have different effects. When considering the role of context on a global scale, it is obvious that an individual with, say, two non-marital partners is at much higher risk of HIV infection when living in Zambia than in England. Even within the smaller scale of a town, context is an important determinant of disease. This analysis showed that neighbourhood SES and market proximity are as important risk factors for HIV infection among young women as sexual activity or syphilis. To confine analysis to individual and behavioural factors is ignoring the underlying causes and modifying factors of behaviour. In studies where there is no information on population-level SES, it might therefore be useful to at least adjust for area.

The interplay of population-level and individual risk factors is complex and probably differs depending on time and place. Nevertheless, it seems important to try and disentangle risk factors at different levels in order to identify those populations and individuals most at risk and to understand the reasons why they are at risk, as this would help to design the most appropriate intervention and prevention measures.

Acknowledgements

Members of the Study Group on Heterogeneity of HIV Epidemics in African Cities: A. Buvé (coordinator), M. Laga, E. Van Dijck, W. Janssens, L. Heyndricks (Institute of Tropical Medicine, Belgium); S. Anagonou (Programme National de Lutte contre le SIDA, Benin); M. Laourou (Institut National de Statistiques et d'Analyses Economiques, Benin); L. Kanhonou (Centre de Recherche en Reproduction Humaine et en Démographie, Benin); E. Akam, M. de Loenzien (Institut de Formation et de Recherche Démographiques, Cameroon); S.-C. Abega (Université Catholique d'Afrique Centrale, Cameroon); L. Zekeng (Programme de Lutte contre le Sida, Cameroon); J. Chege (The Population Council, Kenya), V. Kimani, J. Olenja (University of Nairobi, Kenya); M. Kahindo (National AIDS/STD control programme, Kenya); F. Kaona, R. Musonda, T. Sukwa (Tropical Diseases Research Centre, Zambia); N. Rutenberg (The Population Council, USA); B. Auvert, E. Lagarde (INSERM U88, France); B. Ferry, N. Lydié (Centre français sur la Population et le Développement, France); R. Hayes, L. Morison, H. Weiss, J.R. Glynn (London School of Hygiene & Tropical Medicine, UK); N.J. Robinson (Glaxo-SmithKline R&D, UK, formerly INSERM U88); M. Caraël (UNAIDS, Switzerland).

We thank Felicity Clemens for advice on statistical issues concerning clustered data. The original study was supported by the following organizations: UNAIDS, Geneva, Switzerland; European Commission, Directorate General XII, Brussels, Belgium; Agence Nationale de Recherches sur le SIDA/Ministère français de la coopération, Paris, France; DFID, London, UK; The Rockefeller Foundation, New York, USA; SIDACTION, Paris, France; Fonds voor Wetenschappelijk Onderzoek, Brussels, Belgium; Glaxo Wellcome, London, UK; BADC, Belgium Development Cooperation, Nairobi, Kenya. J. R. Glynn is supported by the UK Department of Health (Public Health Career Scientist award).

References

Aral SO, Padian NS & Holmes KK (2005) Advances in multilevel approaches to understanding the epidemiology and prevention of sexually transmitted infections and HIV: an overview. *Journal of Infectious Diseases* **191**, S1–6.

Bloom SS, Urassa M, Isingo R, Ng'weshemi J & Boerma JT (2002) Community effects on the risk of HIV infection in rural Tanzania. *Sexually Transmitted Infections* **78**, 261–266.

Boerma JT & Weir SS (2005) Integrating demographic and epidemiological approaches to research on HIV/AIDS: the proximate-determinants framework. *Journal of Infectious Diseases* **191**, S61–67.

Braveman PA, Cubbin C, Egerter S *et al.* (2005) Socioeconomic status in health research: one size does not fit all. *JAMA* **294**, 2879–2888.

Buve A, Carael M, Hayes RJ *et al.* (2001a) Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa: methods and prevalence of HIV infection. *Aids* **15**, S5–14.

Buve A, Lagarde E, Carael M *et al.* (2001b) Interpreting sexual behaviour data: validity issues in the multicentre study on factors determining the differential spread of HIV in four African cities. *Aids* **15**, S117–126.

Crampin AC, Glynn JR, Ngwira BM *et al.* (2003) Trends and measurement of HIV prevalence in northern Malawi. *Aids* **17**, 1817–1825.

De Walque D, Nakiyingi-Miiro JS, Busingye J & Whitworth JA (2005) Changing association between schooling levels and HIV-1 infection over 11 years in a rural population cohort in south-west Uganda. *Tropical Medicine and International Health* **10**, 993–1001.

Diez-Roux AV (1998) Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health* **88**, 216–222.

Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Tembo F & Monze M (2001) Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. *Aids* **15**, 907–916.

Glynn JR, Carael M, Auvert B *et al.* (2001) Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia *Aids* **15**, S51–60.

Glynn JR, Carael M, Buve A *et al.* (2004) Does increased general schooling protect against HIV infection? A study in four African cities *Tropical Medicine and International Health* **9**, 4–14.

Gregson S, Waddell H & Chandiwana S (2001) School education and HIV control in sub-Saharan Africa: from discord to harmony? *Journal of International Development* **13**, 467–485.

Hargreaves JR (2002) Socioeconomic status and risk of HIV infection in an urban population in Kenya. *Tropical Medicine and International Health* **7**, 793–802.

Hargreaves JR & Glynn JR (2002) Educational attainment and HIV-1 infection in developing countries: a systematic review. *Tropical Medicine and International Health* **7**, 489–498.

Johnson L & Budlender D (2002) *HIV risk factors: a review of the demographic, socio-economic, biomedical and behavioural determinants of HIV prevalence in South Africa*. CARE monograph, No.8, Centre for Actuarial Research, Cape Town.

Kilian AH, Gregson S, Ndyanabangi B *et al.* (1999) Reductions in risk behaviour provide the most consistent explanation for declining HIV-1 prevalence in Uganda. *Aids* **13**, 391–398.

Konde-Lule JK, Wawer MJ, Sewankambo NK *et al.* (1997) Adolescents, sexual behaviour and HIV-1 in rural Rakai district, Uganda. *Aids* **11**, 791–799.

Laga M, Schwartlander B, Pisani E, Sow PS & Carael M (2001) To stem HIV in Africa, prevent transmission to young women. *Aids* **15**, 931–934.

Lewis J, Donnelly C, Mare P, Mupambireyi Z, Garnett G & Gregson S (2006) *Evaluating the proximate determinants framework for HIV infection in rural Zimbabwe. XVI International AIDS Conference in Toronto, Canada*. Poster MOPE0351.

Michelo C, Sandoy IF & Fylkesnes K (2006) Marked HIV prevalence declines in higher educated young people: evidence from population-based surveys (1995–2003) in Zambia. *Aids* **20**, 1031–1038.

Over M & Piot P (1996) Human immunodeficiency virus infection and other sexually transmitted diseases in developing countries: public health importance and priorities for resource allocation. *Journal of Infectious Diseases* **174**, S162–175.

Pearce N (1996) Traditional epidemiology, modern epidemiology and public health. *American Journal of Public Health* **86**, 678–682.

Rose G (1992) *The Strategy of Preventive Medicine*. Oxford University Press.

Susser M & Susser E (1996) Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *American Journal Public Health* **86**, 674–677.

Victora CG, Huttly SR, Fuchs SC & Olinto MT (1997) The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *International Journal of Epidemiology* **26**, 224–227.

Wojcicki JM (2005) Socioeconomic status as a risk factor for HIV infection in women in East, Central and Southern Africa: a systematic review. *Journal of Biosocial Science* **37**, 1–36.

Zutavern A (2001) *The epidemiology of HIV infection among young women in Kisumu*, Kenya. LSHTM.

Corresponding Author Sabine Gabrysch, London School of Hygiene & Tropical Medicine, 8 Bedford Square, G10, London WC1B 3RA, UK. Tel.: +44 20 7958 8173; E-mail: sabine.gabrysch@lshtm.ac.uk