2

The effects of societal variables on urban rates of HIV infection in developing countries: An exploratory analysis

by Mead Over[1]

Section One: Introduction

One of the continuing puzzles of the AIDS epidemic is the variation from one country to the next in the percentage of the population infected with HIV (the HIV prevalence rate). Differences in the level of infection and in its rate of spread can sometimes be dramatic, with infection rates rising rapidly and tragically in one country and remaining stable at low or almost nonexistent levels in another.

Since HIV is transmitted between adults primarily by unprotected sex, once HIV is introduced among the sexually active population of a country, differences in the rate of spread must be attributable to differences either in sexual behaviour or in the infectivity of the virus. The three major causes of increased infectivity of the virus that have been proposed in the literature are:

- the presence of other 'classic' sexually transmitted diseases like syphilis or gonorrhoea (Grosskurth et al 1995)
- the absence of male circumcision (Bongaarts et al 1989)
- the specific subtype or 'clade' of HIV prevalent in the country.

The first two of these are related to socioeconomic and cultural factors: other sexually transmitted diseases are more prevalent when people have more unprotected sex and when people fail to get prompt treatment for such diseases; male circumcision is largely governed by traditional customs and religious beliefs. The third, the genetic variation in the virus itself, is still speculative and, since the mix of subtypes present in a country is not known for more than a few countries, must remain beyond the scope of this study.

While the determinants of an individual's sexual activity are subtle and complex, it is reasonable to expect that social, cultural, and economic conditions will influence the frequency of risky sexual behaviour. Furthermore, such aggregate-level social variables will also influence the prevalence of other sexually transmitted diseases and the percentage of the male population that is circumcised. Thus, it is plausible to imagine that variations in social and economic variables might predict the level of HIV infection in a given country. The purpose

of this paper is to explore the available data on HIV prevalence in urban areas of developing countries in order to discover whether, and to what degree, available aggregate-level data can explain the observed HIV prevalence rates.

Section Two: The dependent variable and the functional form

The study of cross-country variation in HIV infection rates is greatly facilitated by the US Bureau of the Census' (BUCEN) compilation and dissemination of all data on such rates that has ever been published or presented at an international conference (US Bureau of the Census 1997). Each new release of their database is accompanied by a summary table containing their choice of the most representative recent estimate of HIV infection in each of four broadly defined population groups. The present analysis focuses on two of these:

- urban high-risk groups, which BUCEN defines as prostitutes or STD patients
- urban low-risk groups, defined as members of the general population or as pregnant women attending antenatal clinics.

HIV infection rates vary a great deal from one country to another. Of the 83 countries for which BUCEN has recent data, from 0 to 8.3 per cent of urban groups practising high-risk behaviour are infected. Fourteen countries have less than 1 per cent infection rates in these groups and at the other extreme another 14 have more than 45 per cent infection rates. According to BUCEN data on the infection rate among low-risk groups in 96 countries, infection remains at 0 per cent in 29 of the countries, but ranges up to 33 per cent in others.

The percentage of the population infected by any contagious disease typically follows an S-shaped curve, starting slowly, then accelerating, and then levelling off (World Bank 1997). In order to apply linear multiple-regression analysis to a variable with such a nonlinear underlying structure, we transform the HIV prevalence rate. First, we assume a maximum or 'ceiling' rate for each of the two risk groups that is greater than any rate yet observed: 90 per cent in the high-risk group and 40 per cent in the low-risk group. Let y be the infection rate in a certain risk group in the urban area of a given country and C be the ceiling assumed for that risk group. We define logit (y) for that risk group in that country as:

$$\log_e [y/(C-y)].$$

In order to avoid dropping observations where an infection rate was measured to be zero, low-risk zeros are set to 0.1 per cent and high-risk zeros to 1 per cent. These 'logits' of prevalence rates then become the variables to be explained by the statistical analysis.

Section Three: The explanatory variables

What aggregate variables are available that might plausibly explain the riskiness of sexual behaviour among the urban population of a country? The demand by

men for commercial sex seems likely to be greater in cities where the ratio of males to females in the sexually active age range is greater than one. Furthermore, young men in the military and those who have recently arrived in the country might be less likely to establish permanent relationships with women and therefore be more likely to seek commercial sex. Variables are available to measure the ratio of urban men to women ages 20 to 39, the number of soldiers as a per cent of the urban population, and the number of foreign-born in the country as a percentage of the total population.

The number of people who are willing to supply commercial sex in a given city is likely to be greater when alternative employment is hard to find or poorly paid. Although no reliable measure of urban unemployment or wage rates could be found, the purchasing power parity measure of gross national product (GNP) per capita seems likely to be roughly correlated with the availability of good jobs to low-income people. And for any given level of average GNP per capita, the opportunities available to the poorest will be more attractive in a country where incomes are distributed relatively equally. We use the Gini coefficient as a measure of the inequality of income distribution, where a value of zero indicates perfect equality (i.e., everyone earns the same amount) and a value of one indicates perfect inequality (i.e., all income is received by a single person).

Since those practising commercial sex are more frequently women, a second relevant measure of inequality is the gap between male and female literacy rates. In a city of a given average income and a given distribution of that income between rich and poor, a larger gap between male and female literacy is likely to mean fewer conventional job opportunities or lower conventional wages for women. Incidentally, greater inequality between the rich and the poor or between men and women also seems likely to increase the bargaining power of men in the upper-income brackets who would like to purchase commercial sex and thus to augment the demand for, as well as the supply of, commercial sex.

A detailed understanding of the sociology of sexual behaviour in a given country would no doubt help to explain the willingness of people either to supply or to demand commercial sex. Furthermore, in many countries the epidemic is exacerbated by the frequent unprotected sex acts that occur outside the commercial market as usually defined. In the absence of published data on the cultural norms that shape a society's sexual practices, we use the percentage of the population that is Muslim as an indicator, which we expect to correlate negatively with both the supply of and the demand for risky sexual behaviour.

A final important consideration in explaining the current level of infection in a country is the time since the epidemic started. Although an ideal measure of this variable is not available, an alternative is the number of years since the first case of HIV infection or AIDS was reported. We also

experiment with using regional dummy variables to proxy both the age of the epidemic and other difficult-to-measure social and cultural differences across regions.

Section Four: Regression results

Table 1 presents the results obtained when we regress the logit of infection rate on these eight variables.

- Columns (1) and (2) of the table present the means and standard deviations of the logits of infection rates and of the eight explanatory variables for a total of 72 countries.
- Columns (3) and (4) present the straightforward application of ordinary least squares regression for infection rates among high-risk and low-risk groups, independently.

Table 1. Regressions explaining the percentage of the urban population infected with HIV

			Regression results[b]					
Risk group	**Descriptive statistics[a]**		**High risk**	**Low risk**	**Low risk**	**Pooled**		
Column number	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Subset(s) of countries	A	B	A	A	A	A	A & B	A & B
Age of epidemic	10.90	11.18	0.53	0.42	0.24	0.47	0.63	0.61
	2.0	*1.5*	*3.1*	*2.4*	*1.3*	*3.6*	*5.5*	*5.2*
GNP per capita (log)	7.9	7.2	-0.7	-1.3	-1.0	-1.0	-1.2	-1.2
	0.8	*1.9*	*-1.9*	*-2.9*	*-2.3*	*-2.9*	*-5.2*	*-4.9*
Foreign-born per cent (log)	0.14	0.74	0.20	0.38	0.30	0.28	0.25	0.24
	1.4	*1.2*	*1.1*	*2.0*	*1.6*	*2.0*	*1.9*	*1.8*
Per cent Muslim	24.0	23.7	-0.016	-0.022	-0.017	-0.018	-0.013	-0.014
	36.2	*31.3*	*-2.0*	*-2.6*	*-2.0*	*-3.3*	*-3.2*	*-3.3*
Gini index of inequality	0.4	0.3	9.8	11.4	8.4	10.6	7.9	8.7
	0.1	*0.3*	*3.2*	*3.5*	*2.5*	*4.0*	*3.7*	*3.8*
Male-female literacy gap	15.3	12.2	0.064	0.074	0.051	0.069	0.051	0.059
	10.3	*12.6*	*1.7*	*2.0*	*1.4*	*2.4*	*2.2*	*2.4*
Urban M/F gender ratio age 20–39	1.1	0.9	3.3	2.5	1.5	2.9	2.4	2.2
	0.2	*0.5*	*2.0*	*1.5*	*0.9*	*2.6*	*1.9*	*1.8*
Military forces as a per cent of the urban population	10.9	7.6	0.070	0.042	0.020	0.056	0.056	0.055
	13.1	*7.5*	*3.1*	*1.7*	*0.8*	*3.1*	*3.1*	*3.1*
Constant			-11.3	-7.2	-3.7	-10.2	-8.5	-8.5
			-2.5	*-1.6*	*-0.8*	*-2.6*	*-2.7*	*-2.7*
Logit of high-risk prevalence	-2.1	-1.6			0.33			
	2.2	*2.0*			*2.2*			
Logit of low-risk prevalence	-3.5	-3.0						
	2.4	*1.8*						
High-risk group dummy						1.6	1.5	1.5
						5.9	*6.6*	*6.8*
Number of observations, maximum	50	22	50	46	46	96	132	132

Risk group	Descriptive statistics[a]		Regression results[b] High risk	Low risk	Low risk	Pooled		
Column number	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Number of countries	50	22	50	46	46	50	72	72
Max. missing variables per obs.	0	1	0	0	0	0	1	1
Number of missing values	0	22	0	0	0	0	36	36
Filled by means or fitted values							Means	Fitted
R^2			0.586	0.650	0.691	0.636	0.621	0.625
F-statistic of the regression			7.3	8.6	9.0	23.2	20.5	20.8
Degrees of freedom for F			(8, 41)	(8, 37)	(9, 36)	(9, 49)	(13, 71)	(13, 71)
Root mean squared error			1.5781	1.5794	1.5046	1.5380	1.5132	1.5049
F-test of region dummies (p-value)			0.092	0.036	0.016	0.051	0.052	0.154

a. For columns (1) and (2) the numbers for each variable are the mean and, in italics, the standard deviation. The number of observations is 50 for all variables in column A except the logit of low-risk prevalence, for which only 46 observations are available. In column B, variables have 22 observations, with the following exceptions: GNP per capita (21), per cent Muslim (20), Gini index (12), literacy gap (18), gender ratio (18), military forces as per cent of urban population (21), logit of high-risk prevalence (16), logit of low-risk prevalence (20).

b. For columns (3)–(8), the numbers for each variable are the estimated regression coefficients and, in italics, the t-statistics for those coefficients. In columns (6) – (8) standard errors are corrected for the heteroskedasticity induced by including two observations for each country. The heteroskedasticity corrections are performed by using the 'cluster' option on STATA's 'regress' command. In columns (3)–(6) there are no missing values of any variables. In columns (7) and (8) there is no more than one missing value on any observation, but individual variables can have several missing values according to the pattern described in note a for column (2). Missing values are replaced by their means (7) or by their fitted values (8).

- Column (5) expands the model of column (4) by allowing the (logit of the) current period infection rate among high-risk groups to affect the infection rate among low-risk groups.
- Columns (6) through (8) pool the data from the high- and low-risk groups in order to estimate a single regression equation for both groups.

What lessons can one learn from these regressions? In order to interpret the coefficients, one must keep in mind that the dependent variable is measured in 'logits' as defined above. At the sample mean, a one-unit change in the logit is equivalent to a change of about 4.3 percentage points in HIV infection among low-risk urban adults.

Columns (3) and (4) demonstrate, first of all, that the eight selected societal variables are capable of explaining one-half to two-thirds of the cross-country variation in infection rates. Since cross-country regressions are known for their typically poor explanatory power, these regressions can be said to fit remarkably well. Furthermore, in each regression all coefficients have the expected sign and six of eight have t-statistics equal to 1.9 or larger, indicating that the associated coefficients are statistically significant at the .05 significance level. These fits are obtained without including regional dummy variables in the regressions.

As is indicated in the last row of table 1, the hypothesis that all regional dummies are jointly equal to zero can only be rejected at the .05 significance level in two of the six reported regressions. Therefore, we do not report the estimates that include the regional dummies for any of the regressions.

The regression equations (3) and (4) can be thought of as the 'reduced form' of a two-equation structural model in which each of the two urban infection rates, together with a subset of the societal variables, affects the other. Such models are difficult to specify, identify, and estimate on aggregate data unless one is willing to adopt substantial additional prior assumptions. One plausible simplifying assumption is that the infection rate in the high-risk group affects that in the low-risk group, with no feedback in the other direction. Under this assumption the structural model is said to be 'recursive' and can be estimated without bias by ordinary least squares. Equations (3) and (5) are the estimates of the high- and low-risk equations in the recursive model.

The estimates in column (5) support the recursive model, but not to the exclusion of all other social determinants of low-risk infection. This conclusion follows from the fact that the high-risk infection rate has a statistically significant and sizeable coefficient of 0.3 in column (5), but does not drive all of the other societal variables out of the regression. For the average country, a coefficient of 0.3 approximates the elasticity of low-risk prevalence with respect to high-risk prevalence and implies that an increase of 10 percentage points in the infection rate in the high-risk group (e.g., from 10 per cent to 20 per cent infected) is associated with an increase in infection in the low-risk group of a third of a percentage point (e.g., from 1 to 1.33 per cent infected).

Comparing columns (4) and (5), the variables whose coefficients are most attenuated by the inclusion of the high-risk infection rate are the age of the epidemic, the urban male-female gender ratio, and the size of the military as a proportion of the urban population. The evidence of the regression suggests that, while the other five variables directly affect both low- and high-risk infection rates, these three variables act primarily to raise the infection rate in the high-risk group, which in turn drives up the low-risk infection rate. Since the recursive model requires more assumptions than the reduced form model, we focus on the reduced form results in the rest of this paper.

Another striking feature of regressions (3) and (4) is the similarity of their coefficients. Though collected from distinct population groups within the country, the two different measures of infection rate are similarly affected by the eight societal variables, and this is consistent with their being two distinct measures of the same epidemiological process. This similarity is the motivation for increasing the power of the estimates by pooling the low-risk and high-risk data into the same regression equation, which is done in columns (6) through (8). After allowing for a dummy variable to distinguish between the low- and high-risk measures, a Chow test of the hypothesis that the regressions can be pooled cannot reject that hypothesis, even at a 25 per cent significance level.

Pooling the data on low- and high-risk groups into a single regression improves the precision of the coefficient estimates substantially, even after taking into account the within-country correlation between the residuals from the low- and high-risk groups. Note that the pooled estimates in column (6) are a weighted average of the estimates for the two groups separately and that the t-statistics and F-statistic are higher than they were in columns (3) or (4). All t-statistics now exceed the critical value associated with the 5 per cent significance level.

4.1 Admitting 22 additional countries

The 50 countries included in set A and used to estimate equation (6) include 17 from sub-Saharan Africa, 15 from Latin America and the Caribbean, 14 from Asia (including India and China), and 4 from the Middle East. One can wonder whether the results in column (6) are specific to these countries, which have complete data for the eight independent variables and for one or both of the two prevalence rates. Columns (7) and (8) expand the sample to admit the 22 additional countries (set B), which are missing no more than one of the eight explanatory variables. Fifteen of these are from sub-Saharan Africa, five from Latin America and the Caribbean, and one each from Asia and from the established market economies. Descriptive statistics for the available data on countries in set B are presented in column (2).

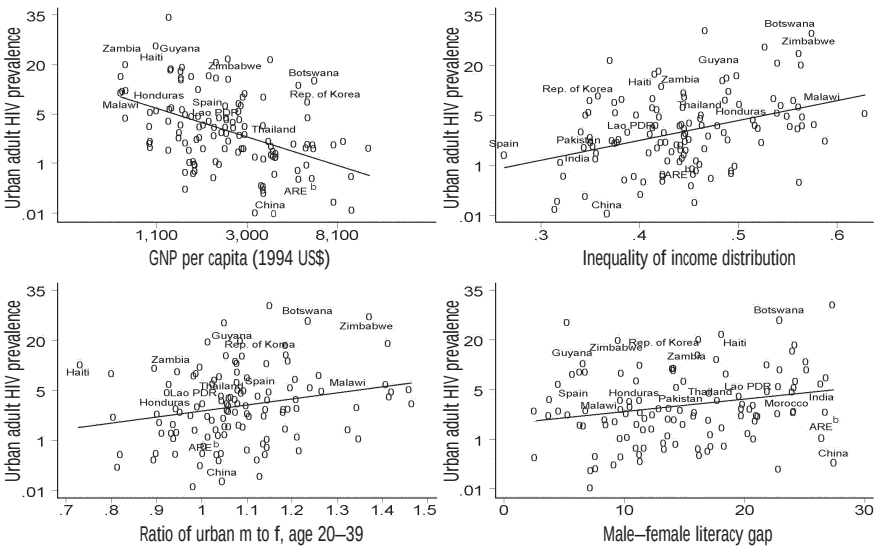
In order to include these 22 additional countries without dropping any of the eight explanatory variables, we impute the missing values of each of the eight variables. Columns (7) and (8) differ only in the imputation method used. In column (7) the missing values are replaced by their means and in column (8) by their predicted values based on a regression in the subset of data for which they are present. Greene (1993) discusses these two approaches under the terms ‘zero-order’ and ‘first-order’ mean replacement. Rather than replace the missing value directly with its imputed value, we zero-fill the missing values and create an auxiliary variable containing the imputed values.

The number of additional variables created in this way equals the number of original explanatory variables that contain at least one missing value. The coefficients of these auxiliary variables are not reported because they are typically very similar to the coefficients of the original variables and their inclusion would make the table unwieldy. However, the number of such variables is reflected in the number of degrees of freedom reported for the F-tests in table 1. (Full results are available from the author.)

Since the estimates in columns (7) and (8) are quite similar, we focus attention on the simpler approach in column (7), in which missing values are replaced by their means. Although the fit of equation (7) is slightly worse than that of (6), we prefer to use equation (7) because its inclusion of a total of 72 countries makes it more representative of the global epidemic.

Figure 1 shows the associations of four of these variables with the percentage of urban adults infected with HIV as estimated in column (7) of

Figure 1. Poverty and gender inequality spread AIDS



Note: The vertical axis measuring HIV infection has been transformed to a logarithmic scale. Points on a given scatter plot represent the data for individual countries after removing the effects of the other seven variables included in the regression analysis. Income inquality is measured by the Gini coefficient. On average, the passage of 1.6 years is associated with an increase of about one logit, or about 4.3 percentage points in the prevalence rate of low-risk urban adults. Source: Column (7) of Table 1.

table 1. Each panel of figure 1 presents the relationship between one of these societal variables and HIV infection after purging that variable and the HIV infection rate of the effects of the other seven explanatory variables. The fitted lines show the estimated slope coefficients from column (7) of table 1. The figures were constructed by the 'avplot' command in the STATA software package (StataCorp 1997). The vertical axis is scaled in logits as defined above.

4.2 Poverty and income inequality

The top two panels of the figure demonstrate that, holding constant the other variables, both low income and unequal distribution of income are strongly associated with high HIV infection rates. For the average country a $2,000 increase in per capita income is associated with a reduction of about 4 percentage points in the HIV infection rate of low-risk urban adults. Although the negative relationship between GNP per capita and HIV infection is strongly supported by this international data, individual-level data do not

always show a negative relationship between individual or household income and HIV infection and frequently demonstrate the converse. Chapter 3 of *Confronting AIDS* (World Bank 1997, 127–32) discusses the contradictory individual-level studies of this topic and possible ways to reconcile them with the aggregate finding reported here. (See also Ainsworth and Semali 1998 and Menon et al 1998 in this volume.)

Reducing the index of inequality from 0.6 to 0.3, the difference in inequality between, say, South Africa and Nepal, is associated with a reduction in the low-risk infection rate by about 3 percentage points. These findings suggest that rapid and fairly distributed economic growth will do much to slow the AIDS epidemic.

4.3 Gender inequality

When examining the influence of gender inequality on HIV infection, one must hold constant as much as possible other influences, which may be correlated with gender inequality across countries. The bottom two panels of figure 1 show that, after controlling for the percentage of the population that is Muslim, per capita GNP, income inequality, and four other societal characteristics, two measures related to gender inequality are associated with higher HIV infection rates.

The first of these, the ratio of males to females in urban centres, varies remarkably across countries, with some countries having fewer male urban residents than female and others having 40 per cent more males. Other things being equal, one might suppose that commercial sex will be more common in cities where men greatly outnumber women and that HIV levels will therefore be higher. The evidence of the lower left panel of figure 1 is that cities in which men ages 20–39 greatly outnumber women do in fact have significantly higher HIV infection rates. For the average country, increasing the job opportunities for young women so that the ratio of males to females in urban areas falls, say, from 1.3 to 0.9, would decrease the low-risk infection rate by about 1 percentage point.

The second measure related to gender inequality included in the analysis is the gap between adult male and female illiteracy rates. Again there is great variation across countries, with the literacy rate among men being as much as 25 percentage points higher than among women in some countries. When women are much less literate than men, they may be less able to bargain effectively with men and thus be at greater risk in sexual encounters. Furthermore, illiterate women will have difficulty finding jobs and thus may depend more on sexual relationships for economic survival, again reducing their bargaining power. The lower right panel of figure 1 supports these ideas, suggesting that the average country that eradicates a 20 percentage point literacy gap between the genders can expect urban low-risk HIV infection to be about 1 percentage point lower.

4.4 Economic growth

From the evidence of figure 1, a country that improves per capita income and reduces inequality, for example, by implementing broad pro-growth investment policies that generate jobs, will reduce its risk of suffering an AIDS epidemic or help to minimise an epidemic already under way. If in addition, the country acts to close the literacy and urban employment gaps between men and women, HIV will have even more difficulty spreading. Unfortunately some of the very processes that might achieve these goals can also stimulate the spread of AIDS. And other policies that sometimes accompany growth, without necessarily contributing to it, can likewise worsen the epidemic.

An open economy is recognised as a key requirement of rapid growth. Openness primarily refers to the ease with which entrepreneurs can move goods and capital across national borders, but a higher degree of openness to trade and financial flows is typically also accompanied by a greater degree of openness to movements of people, including immigration. Moreover, some studies have suggested that migration itself is correlated with growth. This would hardly be surprising, since immigrants are often among the hardest working and most enterprising people in any country.

However, the regression analysis suggests that countries with larger immigrant populations tend to have larger AIDS epidemics: other things being equal, a country in which 5 per cent of the population is foreign-born can expect to have an infection rate about 2 per cent higher than a country with no foreign-born residents.

4.5 Militarisation

A final factor, which in the regression analysis is not necessarily associated with development but can be readily affected by government policy, is the level of militarisation. In developing countries military forces are often based near urban centres and consist almost entirely of unmarried men. Using a variable that measures the number of men in the nation's armed forces as a percentage of its urban population, the regression analysis shows that even after controlling for the ratio of male to female urban residents, countries with more soldiers will have higher infection rates. For the average country, reducing the size of its military from 30 per cent to 12 per cent of the urban population will reduce seroprevalence among low-risk urban adults by about 1 percentage point.

4.6 Other variables

The age of the epidemic is statistically significant in all specifications except, as noted above, in column (5). The size of the column (7) regression coefficient on the age of the epidemic is consistent with an average rate of increase of 0.7 percentage points per year in the infection rate among low-risk groups and about 5.7 percentage points per year in the infection rate among those at high risk.

We have experimented with a variety of other variables that might plausibly have been thought to influence HIV prevalence. Introduced one at a time, none of the following variables contributes to a statistically significant degree to explaining the cross-country variation in HIV prevalence: the ratio of male to female secondary school enrolment, rate of urban population growth, percentage of male population circumcised, the availability of condoms in 1982 and 1990, and public health expenditure as a percentage of gross national product.

Section Five: Conclusion

An exploratory analysis of the determinants of HIV infection rates demonstrates a surprisingly strong relationship between eight aggregate variables and the infection rate of low-risk urban adults. The fact that low-risk and high-risk infection rates display similar relationships with these eight variables leads us to pool the two measures of severity in order to increase the power of the estimates. The results of the analysis are broadly consistent across several different specifications and with or without the inclusion of an additional 22 countries that have missing values on one or more of the eight independent variables. The results support several hypotheses that have been common in the literature on AIDS:

- Both poverty and income inequality facilitate the spread of AIDS.
- Inequality between the sexes, either in access to urban jobs or in literacy, can speed the spread of AIDS.
- Countries with more soldiers have higher HIV infection rates, a finding that points to the importance of condom promotion among the military.
- High-risk groups are the driving force of the epidemic.

Although the variables used to capture openness of the economy and of the frontiers to labour force migration are imperfect, the percentage of the population that is foreign-born has a statistically significant positive effect on infection rates in two of the six reported regressions. This finding suggests that some of the processes that are friendly to growth might also exacerbate the AIDS epidemic, unless public health officials take compensating protective measures.

A path toward control of the AIDS epidemic is economic and social development, implemented so as to reduce, rather than to increase, income and gender inequality. But policies to speed such growth are already desirable for general development reasons. The analysis here points to the need for additional policies to reduce risky behaviour among soldiers, sex workers, and others who are most likely to contract and spread HIV.

Furthermore, the regressions fail to explain fully two-fifths of the variation in infection rates across countries. Some of this additional variation might be explained with better measures of the variables we are already analysing. But most of it is undoubtedly due to subtle characteristics of the cultures of these

countries and of their efforts to combat the AIDS epidemic that will remain difficult or impossible to measure at the aggregate or national level. Factors such as the degree of leadership provided by national figures or the quality of peer education programmes among sex workers will ultimately have an enormous impact on the success of national HIV control efforts. Until an HIV vaccine becomes available, effective control policy will require a combination of fair, pro-growth macroeconomic policies with multi-pronged, sector-specific, tightly focused interventions that ensure reductions in risky behaviour of those most likely to contract and spread HIV infection, while protecting them from discrimination.

Note

1 The analysis reported here builds on earlier work by Over and Piot (1993) and benefits from the author's access to an incomplete draft on the same topic by David Bloom and Ajay Mahal (1997). The paper could not have been written without the research assistance of Samantha Forusz.

References

Ainsworth, Martha, and Innocent Semali. 1998. 'Who is most likely to die of AIDS? Socioeconomic correlates of adult deaths in Kagera Region, Tanzania'. This volume.

Bloom, David E., and Ajay S. Mahal. 1997. 'The AIDS epidemic and economic policy analysis.' In David E. Bloom and Peter Godwin, eds. The Economics of HIV and AIDS: The Case of South and South East Asia. Delhi: Oxford University Press.

Bongaarts, John, P. Reining, Peter Way, and Francis Conat. 1989. 'The relationship between male circumcision and HIV infection in African populations.' AIDS 3 (6): 373–77.

Greene, William H. 1993. Econometric Analysis. 2d ed. New York: Macmillan Publishing.

Grosskurth, H., F. Mosha, J. Todd, K. Senkoro, J. Newell, A. Klokke, J. Changalucha, B. West, P. Mayaud, A. Gavyole, R. Gabone, D. Mabey, and R. Hayes. 1995. 'A community trial of the impact of improved sexually transmitted disease: The HIV epidemic in rural Tanzania – baseline survey results'. AIDS 9 (8): 927–34.

Menon, Rekha, Maria J. Wawer, Joseph K. Konde-Lule, Nelson K. Sewankambo, and Chuanjun Li. 1998. 'The economic impact of adult mortality on households in Rakai district, Uganda'. This volume.

Over, Mead, and Peter Piot. 1993. 'HIV infection and sexually transmitted diseases.' In Dean T. Jamison, W. Henry Mosley, Anthony R. Measham, and José Luis Bobadilla, eds. Disease Control Priorities in Developing Countries. New York: Oxford University Press.

StataCorp. 1997. Stata Statistical Software: Release 5.0. College Station, TX: Stata Corporation.

US Bureau of the Census. 1996. World Population Profile 1996, with a special chapter focusing on adolescent fertility in the developing world. Washington, D.C.: US Department of Commerce, US Government Printing Office.

—. 1997. 'Recent HIV seroprevalence levels by country: January 1997'. Research Note 23. Health Studies Branch, International Programs Center, Population Division. US Bureau of the Census, Washington, D.C.

US Bureau of the Census (database). 1997. 'HIV/AIDS surveillance database'. Version 1.1, release 0. Washington, D.C.: Population Division, International Programs Center.

World Bank. 1997. Confronting AIDS: Public Priorities in a Global Epidemic. New York: Oxford University Press.