

Migration, mines and mores: the HIV epidemic in southern Africa[†]

John Hargrove*

The seriousness of the HIV epidemic in southern and eastern Africa has its roots in the 19th century—in the employment practices instituted on mines, farms and in cities, where millions of men have, ever since, lived apart from their families for the greater part of each year. This destruction of the family unit was a sociological disaster waiting for the arrival of HIV and is the source of many other social ills—not least the increasingly violent nature of South African society. In the short term we can promote HIV prevention measures such as male circumcision and condom use. In the medium term, we can hope that the many billions already spent on microbicide and vaccine research begin to pay dividends. In the long term, we need to change fundamentally the way that people live.

Introduction

In a lecture given recently at the South African Centre for Epidemiological Modelling and Analysis (SACEMA) in Stellenbosch, Brian Williams of the World Health Organization (WHO) worried that, 30 years into the HIV/AIDS epidemic, we still do not have a convincing explanation of why that epidemic is so much more severe in some parts of the world than in others. He noted: 'Circumcision is thought to be important—but men in

*This article is based on the author's inaugural lecture delivered on 14 November 2007 at the University of Stellenbosch.

[†]SACEMA, The DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis, University of Stellenbosch, 19 Jonkershoekweg, Stellenbosch 7600, South Africa. E-mail: jhargrove@sun.ac.za

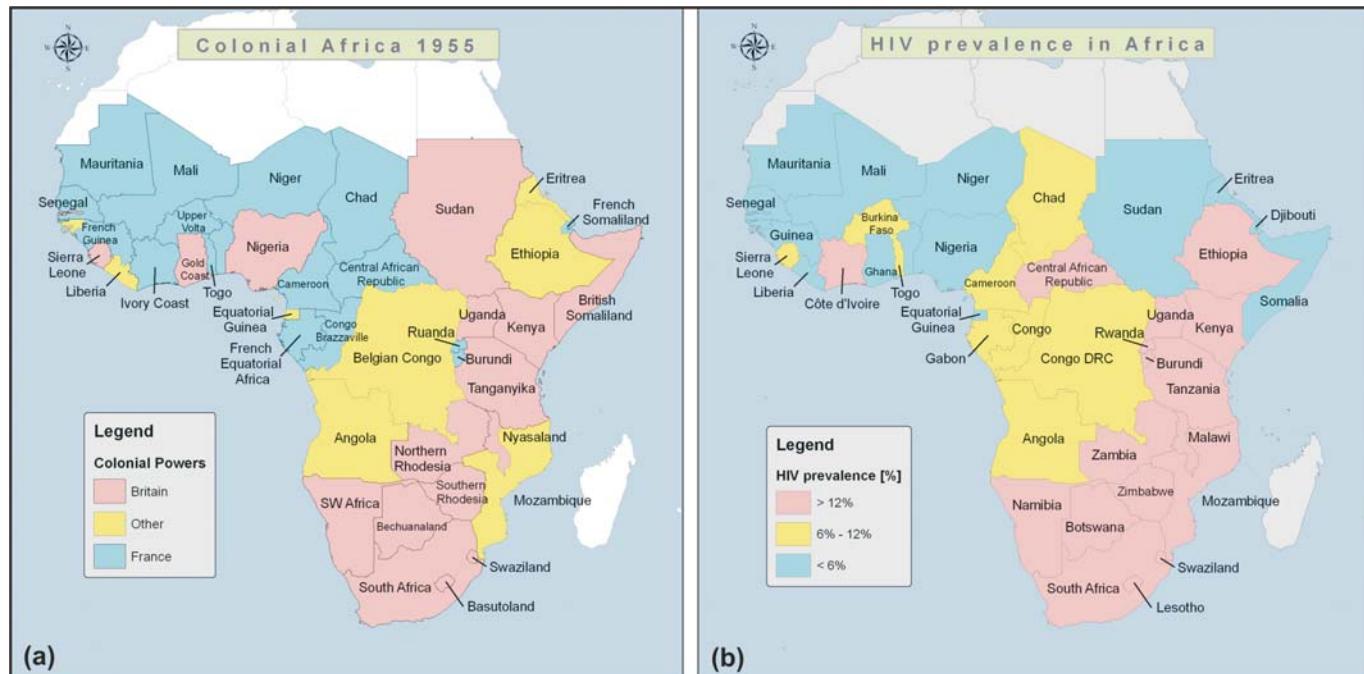


Fig. 1. a, Map of sub-Saharan Africa by colonial power in 1955. b, Sub-Saharan Africa by distribution of peak HIV prevalence in urban areas, up to 2007. Source: UNAIDS Fact Sheets.

North India are not generally circumcised. Promiscuity is thought to be important—but Brazilians are not reticent about sex, and people in West Africa have more sex but less HIV than people in East Africa. Sex before marriage is common in both the US and Europe—but those populations also have low rates of HIV. Women's empowerment is reckoned to be important—but Islamic countries have almost no HIV.'

In illustrating the point, Williams showed a map of the distribution of HIV in Africa. It was a map that took me back more than 50 years to a schoolroom in Bulawayo in 1955 in what was then Southern Rhodesia. In those days, we saw the world through pink-tinted glasses. Almost everywhere we looked on the map we saw the cheerful pink of colonial British Africa (Fig. 1a): the Union of South Africa, Bechuanaland, South West Africa, Southern and Northern Rhodesia, Nyasaland, Basutoland, Swaziland and, further north, Tanganyika, Kenya and Uganda. Fifty years on, Britain has left and a lot of the names have changed. But we still see bright pink all around us (Fig. 1b), though now for less cheerful reasons. Today, we see the map of southern Africa, coloured pink, because shades of pink are used by the WHO and UNAIDS to denote levels of HIV infection, with the deepest shades reserved for those countries with the highest prevalence, the most severe epidemics. And the same eleven countries in southern and east Africa remain pink because they have the dubious distinction of constituting eleven of the top fifteen countries in the world in terms of the level at which adult HIV prevalence, in urban centres, has peaked in their individual epidemics.

The four other countries—Mozambique, Rwanda, Burundi and Ethiopia—are not ex-British colonies, but are also all in

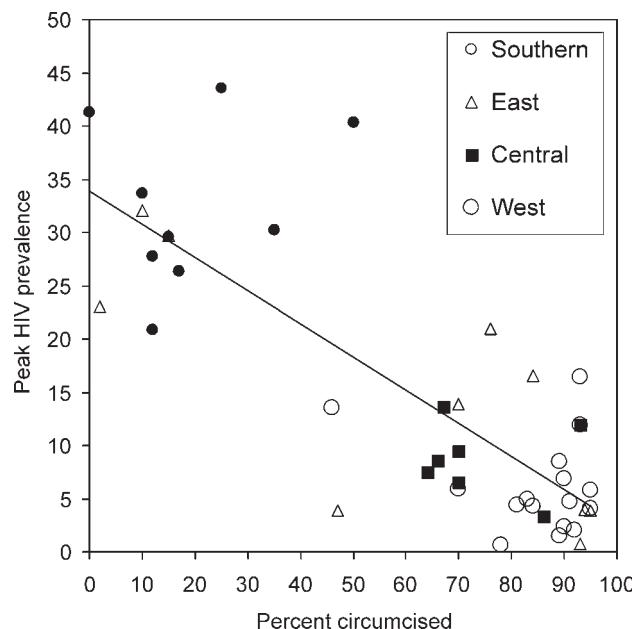


Fig. 2. Peak HIV prevalence as a function of the proportion of men circumcised in the countries of four different regions of sub-Saharan Africa.

southern or east Africa. Although the images in Fig. 1 are very striking, and while former British colonies have an average peak HIV prevalence level of 11% higher than the countries with a different colonial past, peak prevalence seems to be more strongly related to geographical region. A simple analysis of variance between countries defined as being southern, east, central or west African accounts for 66% of the variance in peak HIV prevalence (Table 1). Countries of southern Africa have, on average, a peak prevalence which is nearly 20% higher than in the states of east Africa, and nearly 25% higher than in west and central Africa.

I attempt in this article to shed some light on why the HIV pandemic should be so particularly severe in southern Africa, and then consider ideas about what we should try to do to rectify this situation—in the short, medium and long term.

Others have carried out ecological analyses of the pan-African distribution of HIV prevalence.¹⁻⁴ It has been shown repeatedly, for instance, that HIV prevalence declines significantly with the proportion of males circumcised in a country. Indeed, much of the recent analysis of the distribution of HIV prevalence in Africa has centred on adjusting HIV prevalence for various other possible risk factors, in order to demonstrate that having made all of these adjustments, there is still a strong significant protective effect of male circumcision.

By itself, the proportion of males circumcised accounts for about 66% of the variance in HIV prevalence for urban sites in 42 sub-Saharan African countries (Fig. 2).^{5,6} The situation is not nearly as clear, however, when analysis is carried out on countries

Table 1. Analysis of variance of peak HIV prevalence by country in four regions of sub-Saharan Africa; $r^2 = 0.66$.

Factor	Coefficient (95% CI; P)
(Constant)	6.4 (2.4, 9.8; 0.002)
Region	
West	[Reference]
Central	2.6 (-4.1, 9.3; 0.44)
East	8.7 (2.8, 14.7; 0.005)
Southern	26.5 (20.3, 32.7; <0.001)

CI, confidence interval.

Table 2. The relationship between peak HIV prevalence and the proportion of Muslims and of males circumcised by country in various regions of sub-Saharan Africa. With these units, for every 1% increase in the fraction that are Muslim (or males that are circumcised) the prevalence decreases by 1% of the value of the corresponding coefficient. In southern Africa, for example, the decrease would be 0.372% per unit increase in the proportion that are Muslim.

Region	Fraction Muslim	Fraction circumcised
	Coefficient (95% CI; r^2 ; P)	
<i>Univariate analyses</i>		
Southern	-37.2 (-74.1, -0.4; 0.37; 0.048)	-8.6 (-55.5, 38.4; 0.03; 0.68)
East	-31.3 (-46.3, -16.2; 0.71; 0.001)	-29.2 (-42.8, -15.6; 0.72; 0.001)
Central/West	-7.4 (-12.1, -2.7; 0.31; 0.003)	-12.8 (-25.8, 0.1; 0.17; 0.052)

within the different regions. Univariate analysis within three African regions produces a significant relationship between HIV prevalence and circumcision rates for east Africa only (Table 2). By contrast, the proportion of Muslims in a country is a significant factor in each region separately.

We should be clear, however, that the last two years have seen the publication of results from three randomized control trials on the effect of medical male circumcision on the acquisition of HIV infection.⁷⁻⁹ All agreed in showing a significant protective effect for men. Medical male circumcision in each case resulted in at least a halving of the annual rate of new HIV infections. Thus, while ecological studies of the distribution of HIV prevalence in Africa may be open to criticism, the incidence studies provide much more compelling and unequivocal evidence.

The protective effect of male circumcision against the transmission of HIV infection from females to males is thus a less contentious issue than it has been, and the results of the ecological studies referred to above are therefore of less interest in this regard. Indeed, given the results in Tables 1 and 2, there is a case for standing the argument on its head. Instead of making adjustments in efforts to demonstrate the protective effects of male circumcision, we could rather adjust for male circumcision and look at the residual variance due to other factors. As we will see, what emerges from this analysis are strong residual regional and religious effects, which suggest what may be the important underlying determinants of the HIV epidemic.

Comparing HIV prevalence levels between countries

In analysing the distribution of HIV prevalence among countries, it has been the practice to compare prevalences at a single point in time.¹⁻⁴ This may be justified on the grounds that standards of HIV testing have changed, and have improved, over the years. It is thus difficult to know how to compare, with time, prevalence estimates from the same country. Regardless of the problems associated with past estimates of HIV prevalence, however, it is not valid to compare HIV prevalence between countries which are at different stages of the epidemic.

A single example should suffice to make this point. HIV prevalence in urban Uganda in the year 2000 was about 8%. In Swaziland it was nearly five times this level, at 37%. But this huge difference is entirely due to the timing of the analyses. If, for example, the comparison had been made in 1992, the Swazi HIV prevalence would have been only about 5%, and the Uganda prevalence six times as high, at nearly 30%.

Given these considerations, it is clear that analysis of HIV prevalence data, and identification of risk factors using multi-country data, should be carried out using data only from a similar stage of the epidemic in each country. It makes sense, in particular, to compare the peak levels of HIV. We obtained all available antenatal clinic data on HIV prevalence in urban settings from 42 sub-Saharan countries, and fitted the time

Table 3. Factors associated with peak HIV prevalence in various countries of sub-Saharan Africa. Multiple regression using Equation (1).

Factor	Univariate Coefficient (95% CI; P)	Multivariate Coefficient (95% CI; P)
a_0 Constant	24.2 (17.0, 31.4; <0.001)	
a_1 Proportion Muslim	-23.9 (-33.6, -14.2; 0.006)	-9.3 (-15.7, -2.9; 0.006)
a_2 Proportion circumcised	-35.4 (-42.1, -28.7; 0.001)	-16.0 (-25.5, -6.6; 0.001)
Region		
a_3 West/Central	[Reference]	[Reference]
East	7.9 (2.4, -13.5; 0.006)	5.3 (0.9, 9.7; 0.020)
Southern	25.7 (19.9, -31.5; <0.001)	13.8 (7.5, 20.0; <0.001)

series of data using a double logistic function.⁵ We attempted to ameliorate the comparability of these estimates over time, and to reduce the effect of serious outliers in the data, by using for analysis the smaller of the observed and predicted peak values of prevalence.

Estimating the effects of male circumcision and religious customs in the same model

HIV prevalence in 122 developing countries has in the past been analysed separately for the effects of male circumcision and religion, on the grounds that the percentage of males circumcised was collinearly correlated with the proportion of Muslims in a population.¹ It was thus impossible to disentangle the proportion of the protective effect accruing to male circumcision from that attributable to other cultural norms particular to the Muslim faith. The effects of religion and male circumcision were separated in later work by examining HIV prevalence in countries stratified on the basis of the proportion of Muslims (or Christians) in the country.² No quantitative measure was, however, afforded of the protective effect of Muslim behavioural norms other than male circumcision.

In another study a similar analysis noted that the impact of circumcision on the AIDS rate was greater than the influence of the Muslim variable, which included circumcision as well as behavioural norms, apparently implying that no added protection accrues from Muslim behavioural norms.³ The following argument shows that this is not the case.

We define: γ^m = fraction of Muslim males circumcised = 1; γ^n = fraction of non-Muslim males circumcised; $\gamma^{m,n}$ = overall fraction of males circumcised; and μ = proportion of Muslims in the population. We initially set up the model for HIV prevalence (π) as the sum of the effects of the proportion that is Muslim, the proportion of males circumcised, and the region:

$$\pi = a_0 + a_1\mu + a_2\gamma^{m,n} + a_3(R), \quad (1)$$

where a_0 is the intercept in the regression model; a_1 and a_2 are the coefficients giving the decrease in HIV prevalence per unit increase in the percentage Muslim, and percentage of males circumcised in the population, respectively; a_3 gives the increase in prevalence in a given region, relative to the region with the lowest prevalence.

Note that for an entirely non-Muslim country (that is, when $\mu = 0$):

$$\pi(0) = a_0 + a_2\gamma^n + a_3(R), \quad (2)$$

from which it is clear that a_2 is the coefficient for the protective effect of being circumcised. We assume implicitly that the protective effect of circumcision is the same for a Muslim as for a non-Muslim man.

In a country which is entirely Muslim (that is, $\mu = 1$ and $\gamma = 1$):

$$\pi(1) = a_0 + a_1 + a_2 + a_3(R), \quad (3)$$

so that the difference in predicted peak prevalence between two

countries, one entirely Muslim and one entirely non-Muslim, but where all males are circumcised (that is, where $\gamma^n = 1$) is given by $\pi(1) - \pi(0) = a_1$. Thus, the total protective effect of being Muslim (as a result of being circumcised and of other cultural factors) is given by $a_1 + a_2$.

Since the model in Equation (1) refers specifically to HIV prevalence among the male population, we should ideally be fitting the model to observed levels of HIV prevalence among males. Reliable data on HIV prevalence among males are unfortunately not available at levels sufficient to support a precise analysis. It is reasonable to assume, however, that the peak HIV prevalence among women in various countries will be highly correlated with the peak prevalence among women in general, and among pregnant women in particular. Accordingly, we apply the above linear model to all available antenatal clinic HIV data from 42 sub-Saharan African countries. Availability of reliable data has also suggested the use of data from urban settings only in each country.

Results and interpretation of the multiple regression analysis

All of the factors involved had highly significant effects on peak HIV prevalence (Table 3), and the inclusion of just three factors accounted for 85% of the variance in the data—despite the inherent variability in, and uncertainty surrounding, the HIV prevalence estimates, and the very simple model used to fit the time series for each country.

Income inequality, as measured by the Gini coefficient, has been cited as playing an important part in determining the variability in the prevalence of HIV in Africa.⁴ In 28 sub-Saharan African countries, for which Gini coefficients were available, peak prevalence of HIV was significantly associated with the Gini coefficient by univariate analysis ($R^2 = 0.28$, $P = 0.004$). However, there were stronger associations with the percentage of Muslims ($R^2 = 0.63$, $P < 0.001$) and the proportion of circumcised males ($R^2 = 0.77$, $P < 0.001$) in a country. In bivariate analyses, using the Gini coefficient and either of these factors, the coefficient was no longer significant ($P > 0.1$ in both analyses). It is thus hard to argue convincingly on currently available evidence for a causal link between a country's HIV prevalence and the degree of income inequality.

Interpretation of the analyses

The results are rather straightforward to interpret, when applying Equation (1) to the results from Table 3. The protective effect of being Muslim can be viewed, crudely speaking, as the sum of protection due to male circumcision ($a_2 = 16.0$), which we assume applies equally to males of all religious persuasion, and protection as a result of other effects ($a_1 = 9.3$) associated particularly with being Muslim. We should thus predict that an entirely non-Muslim country, and one where no male is circumcised, would have a peak HIV prevalence of $16.0 + 9.3 = 25.3$

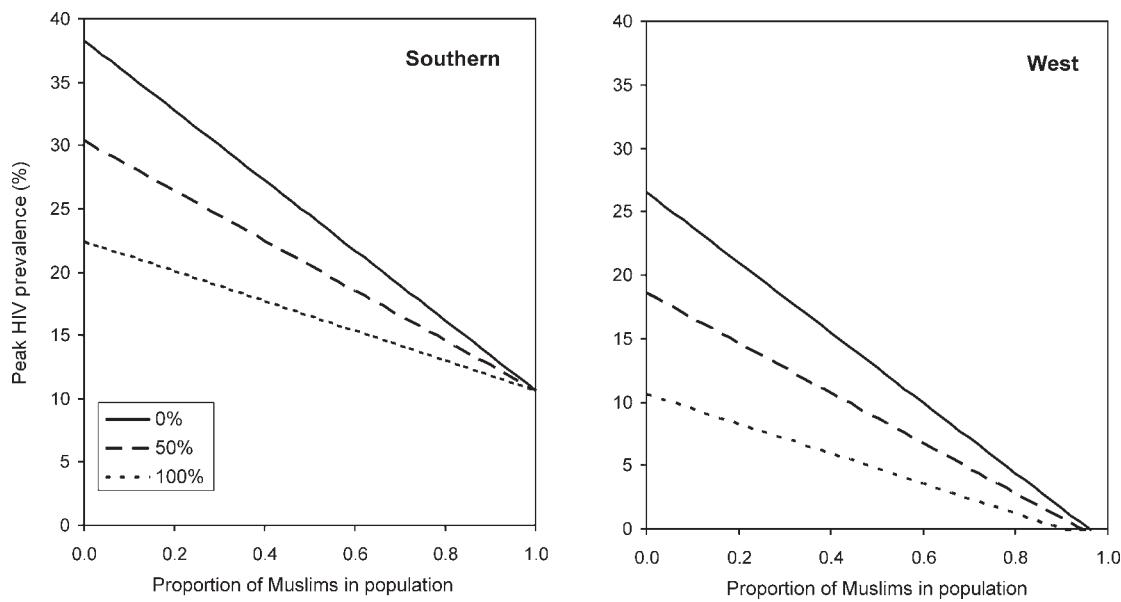


Fig. 3. Peak prevalence as a function of the proportion of Muslims in a country, predicted for various levels of circumcision of non-Muslim males, in two regions of sub-Saharan Africa. Predicted values are approximations only. Thus, predicted levels of zero (and negative) prevalence for proportion of Muslims close to unity simply reflect the joint effects of data variability and violations of the assumptions inherent in the linear model.

percentage points higher than a Muslim country where all males are, by assumption, circumcised. Note that, as in an earlier study,³ we also see in the univariate analysis (Table 3) that the coefficient for the fraction of male population circumcised is larger than that for the fraction of the population that is Muslim. But, as we saw above from the multivariate analysis, this does not mean that there is no added protective effect as a result of being Muslim.

Figure 3 depicts the effects in graphic form, and also illustrates the substantial regional differences in HIV prevalence, when due allowance is made for the proportion of Muslims in the population and of the fraction of men who are circumcised. For the nine countries in the southern region, HIV has peaked at 13.8 percentage points higher than in west and central Africa and at 8.5 points higher than in east Africa.

To summarize, ecological studies in the past have tended to concentrate on the importance of male circumcision in reducing the risk of HIV infection. There has been a tendency to try to explain regional differences in HIV prevalence rates in terms of male circumcision rates, inferring that countries in southern Africa which have lower rates of male circumcision than in west Africa also have much higher HIV prevalence as a consequence.

The message from Table 3 is rather different. What we see from this analysis is that even when we adjust for male circumcision rates, and indeed for the proportion of Muslims in various countries, we still find that HIV prevalence in southern Africa is much higher than in the rest, particularly the west, of Africa. What the analysis also indicates is that even after adjustment for male circumcision rates, the protective effect of being Muslim is nearly 60% of that due to male circumcision and does not differ significantly from it.

These two results raise the obvious question: Is there some common factor that we can see, which might simultaneously explain why the risk of HIV infection is so much lower in Muslim communities and so much higher among all countries in southern and, to a lesser extent, east Africa?

Many factors are considered in the ecological studies as possible risk factors for HIV infection and, at least in univariate analyses, many statistically significant effects have been found—population density; female literacy; poverty; skewed wealth distribution; the proportion of young people in a country; human development

index; sexual behaviour patterns such as divorce and polygamy; fertility rates and age at first intercourse; co-infection with other sexually transmitted diseases and tuberculosis; and a large number of variables related to health and health services, immunisation rates, the numbers of doctors in a country, and various measures of sexual behaviour.^{1,2}

But, by multivariate modelling, the only factors among those listed above that had statistically significant effects were: i) the proportion of young people in the population, positively correlated with HIV prevalence; ii) immunisation rates, female literacy levels, and numbers of doctors per head of population—which all correlated negatively with prevalence. None of these factors seems to fit the requirement of being simultaneously responsible for particularly raised HIV prevalence in southern Africa and the reverse in predominantly Muslim countries.

Other factors of potential importance have, however, been ignored in the cited analyses. In a seldom quoted paper in 1991, David Sanders and Abdulrahman Sambo argued that the migrant system of labour in southern and east Africa contributes to family separation and the spread of disease (including AIDS), from urban to rural areas as well as in the opposite direction.¹¹ There is no consideration in the aforementioned ecological studies of such effects. This is not entirely surprising, since data on relative levels of family coherence and of urban/rural movements are difficult to quantify and to compare among African countries. On the other hand, family coherence seems to be a feature that is very strong in Muslim culture, and one that has also been particularly severely compromised in southern Africa. Bearing this in mind, the results encapsulated in Table 3, to wit, family structure, and factors which impact on that structure, may provide important keys to an understanding of the distribution of HIV infection in Africa. It is therefore worth considering some details of the historical development of southern African society, the consequent effects on family structure, and how these may be related to the HIV pandemic.

Rhodes not roads

When considering family separation and migration in southern Africa the word 'apartheid' perhaps comes to mind—apartheid is a favourite whipping boy in discussions of the roots of all evil in southern Africa in general, and for the HIV epidemic in

particular.¹² There can be no doubt that apartheid contributed greatly to the breakdown and fragmentation of social and family life in South Africa, but I think that the problems have much older and somewhat different origins, as I shall now explain.

When one thinks of the role that the movement of people has played in the spread of HIV, there was in the past, and sometimes still is, a tendency to demonize truck drivers.¹³ Parallels were drawn between good road systems and the rapid spread of the virus via commercial sex workers at truck stops. Roads and road transport undoubtedly played a role, but the analysis above identifies a quite different villain: it's not *roads* that we should be thinking of, but *Rhodes*, Cecil John Rhodes. It was Rhodes, Alfred Beit and their mining associates, aided and abetted by Lord Alfred Milner, who sowed the seeds of the southern African HIV epidemic—and many other problems that are essentially sociological in their origin and nature. Let me elaborate.

Following the defeat of British forces at Majuba in 1884, Paul Kruger, as president of the Transvaal Republic, signed, under duress, the Convention of London, which gave the Transvaal independence in all matters except control over its foreign affairs. What neither side was to know was that gold would be discovered in the heart of this quasi-independent territory just two years later. There was gold in unprecedented quantities, and it was gold which Kruger's *volk* were not then in a position to mine and exploit fully for themselves. What then developed is well known. I note only that, according to one analysis, it was essentially Milner's efforts, assisted particularly by funding from Beit, that heaped unbearable diplomatic, and then military, pressure on Kruger—making demands that Milner prayed that Kruger would reject.¹⁴ Ultimately, the pressure told on Kruger, who appeared to have been persuaded by Jan Smuts that, if it had to be war, then the Transvaal should strike first. The Transvaal sent the British an ultimatum and, when it expired, they invaded Natal. Milner's prayers had been answered—though the less-than-edifying period in British colonial history that followed was perhaps not quite what Milner had in mind.

For Milner, and Rhodes, the battle had always been about 'Empire'; but the real impetus behind the battle in South Africa was the gold. Putting aside the rights and wrongs of the Anglo-Boer war, the subsequent explosive expansion of the goldfields led to massive development in South Africa, and further finance for commercial expansion into the neighbouring territories, funded in large part by people such as Rhodes and Beit.

It wasn't just the material wealth that Rhodes and company needed from the neighbouring countries. There was a major problem with the gold on the Rand. The geologists quickly found that most of the gold was deep underground, embedded as tiny speckles in huge volumes of hard quartz matrix (only 10 g of gold in 1 tonne of ore). The deeper the gold, the more expensive it was to mine. All costs had to be kept to a minimum; and labour costs, in particular, had to be cut to the bone, especially because the demand for labour was huge. As a consequence, the South African mines needed manpower from neighbouring countries. And I do mean 'man'; women were almost entirely excluded. There could be no question of having black families living on the Rand; only the males could do the heavy work underground, so only men would be housed.

And housed they were; hundreds of thousands of men at any one time all over the Transvaal, and then the Orange Free State. Not just from South Africa, but from Southern Rhodesia, Northern Rhodesia, Nyasaland, Mozambique, South West Africa, Swaziland and Basutoland. Over the years companies such as WENELA (the Witwatersrand Native Labour Association) brought millions of healthy, strong and virile young men

for a stint on the mines. Their wives and womenfolk were at home all over southern Africa; they saw each other once a year. In between times, well, I don't have to tell you what strong, virile young men are wont to do—particularly when they have rather more money than their peers, in general, and young women in particular. And what of their wives and girlfriends left at home for months at a time? It was a catastrophe waiting for something like HIV to come along.

We should not single out only the mines, and certainly not just the South African mines, as the sole villain of the piece. The British tried to use forced labour in Uganda, and later recruited Rwandans and Malawians to work there on the coffee and tea estates.¹⁵ And the principle of having only males living and working on mines, as well as on farms and plantations—and even as domestic workers in big cities—became the norm all over British southern and, to an extent, eastern Africa. The pattern was more or less similar everywhere, from the Cape through the Rhodesias, Tanganyika and to the 'White Highlands' of Kenya. In cities in Southern Rhodesia, until well into the 1960s, it was illegal for a (white) employer to allow his (black) manservant to have his wife living on the premises with him.

The epidemiological consequences of oscillatory migration patterns

What became entrenched was a system where men typically lived and worked in cities or on large estates. They returned, perhaps only once or twice a year, to their wives and families, who typically had to maintain a foothold in the countryside, farming a small piece of land, with little support.

While there is little, if any, direct evidence to link this system of oscillating migration and the resulting damage to family structure to its epidemiological consequences, we can offer some indirect evidence. Using a diffusion model for the spread of an infection like HIV, it is easy to demonstrate the intuitively obvious fact that even high levels of sexual activity, and partner change, lead to surprisingly low rates of epidemic expansion—as long as all of this sexual activity is also strictly localized. An African version of *Peyton Place*, or *Desperate Housewives*, if you prefer, leads to very high rates of HIV infection—but the focus of infection is highly localized. While concurrent relationships may well be important requisite conditions for building an epidemic, they are not sufficient.¹⁶ What we also need is that the partners, concurrent or not, are geographically separated.

Modelling shows that the spread of the epidemic is spectacularly different if even 10% of a person's sexual activity is directed in a geographically distant location. This model, which mimics the essential features of oscillating migration, also suggests why such migration is an essential component for an epidemic to be as severe as the one we now witness in southern Africa.

There are indications, even within the southern African HIV epidemiological epicentre, that those who have the strongest family structures and psychosocial support systems are also least likely to acquire HIV infection. Thus, multivariate analysis of HIV prevalence data from the ZVITAMBO Project in Harare, Zimbabwe (Table 4), shows that women who are HIV-infected are more likely to be unmarried than HIV-negative women. Moreover, they will not have completed secondary school education, they will rent or live with extended family rather than own their home, be unemployed, have a husband with fewer years of schooling and a non-professional occupation, have an extremely low family income, and no religious affiliation.¹⁶

Low income *per se* is less important than other factors, being associated only with increased HIV-1 prevalence among women with the lowest 3% of incomes. In short, what seems to be impor-

Table 4. Regression analyses of risk factors for prevalent HIV infection among postpartum women enrolled in the ZVITAMBO trial in Zimbabwe (adjusted for date of recruitment, age and parity).

Variable		n	Odds ratio (95% confidence interval)
Marital status	Married	13 198	1
	Single	638	1.36 (1.13, 1.63)
	Divorced	174	1.54 (1.13, 2.11)
	Widowed	100	3.97 (2.53, 6.23)
Mother's religion	Non-Catholic	10 643	1
	Catholic	2 709	1.19 (1.09, 1.31)
	None	705	1.31 (1.11, 1.55)
Mother's education	Secondary	11 515	1
	<Secondary	2 542	1.58 (0.99, 2.53)
Father's education	Secondary	12 701	1
	<Secondary	1 356	2.27 (1.25, 4.13)
Father's occupation	Profession	2 306	1
	Other	11 751	1.22 (1.10, 1.35)
Mother's housing	Own	1 464	1
	Other	12 593	0.43 (0.24, 0.77)
Family income (US\$/month)	≥18	10 332	1
	<18	357	1.45 (1.16, 1.82)
	Unknown	3 378	1.18 (1.08, 1.29)

tant is the strong 'psychosocial support' typical of a settled family lifestyle.

The long-term solution

The preceding analysis strongly supports the concept that the HIV pandemic in this part of the world has its roots in structural deficiencies—and not in the inherently aberrant sexual behaviour wrongly attributed to African populations.¹¹ The paper by Sanders and Sambo,¹¹ published in 1991, was remarkable in that the HIV prevalence rates at that time in the urban areas of Botswana, Lesotho, Swaziland and South Africa were 'only' 8%, 5%, 2% and 4%, respectively. How much more powerfully can the point about structural deficiencies be made today, when the prevalence levels are six to 15 times higher in these southern African countries.⁵ Sanders and Sambo did not explicitly predict the tsunami of HIV that was about to overwhelm South Africa and its immediate neighbours south of the Limpopo. Given that they identified so accurately the structural deficiencies of self-evident importance in the development of the epidemic, the prediction was almost inescapable for those countries that have been the most severely affected by the oscillating migration associated with the mines, farms and cities of southern Africa.

The unavoidable conclusion must be that we need urgently to address the issue of how we can rebuild family structures in southern and eastern Africa. Without such change, we may confidently predict extreme difficulty in dealing effectively not only with the HIV epidemic, but also with many other problems having similar sociological determinants.

Changing HIV prevalence and incidence in Zimbabwe

One must conclude that sociological improvement will be slow in coming, given the extreme difficulty involved in changing society's habits and traditions, particularly the employment practices that have been entrenched for more than a century—and practices that are implicitly so beneficial for big business (regardless of the language spoken by the bosses or the colour of their skins). The essential structural elements that predispose southern Africa to disastrous levels of HIV infection and other social ills are therefore likely to remain largely intact for the foreseeable future. Faced with this reality, can there be any ray of light, any glimmer of hope, in what is now the epicentre of the pandemic? Curiously, I think there is, even within the very heart

of darkness. Despite all the economic, social and legal shambles, and the human-rights abuses that characterize Robert Mugabe's Zimbabwe,¹⁷ we see the curious fact that HIV prevalence among women attending antenatal clinics (ANCs) in Harare has been declining at nearly 10% per annum since about 1998 (Fig. 4).¹⁸ Moreover, basic epidemiological principles suggest that HIV incidence—that is, the rate of occurrence of new infections—started to decline in 1994. Mathematical modelling suggests that these declines could not be attributed solely to the direct effects of increased deaths among HIV-positive individuals, but have been associated with declines in HIV incidence and changes in sexual behaviour.¹⁹

The model, fitted to Zimbabwean ANC data for 2000–2004, suggests that risky sexual behaviour has halved since 2000, though the authors do not suggest why, or how, such behaviour change might have arisen.¹⁹ It appears from more extensive Harare ANC data, however, that HIV prevalence among younger (less than 24-year-old) women has actually been declining since 1994. Since the vast majority of the HIV-positive members of this age group must have been infected relatively recently, this suggests that the decline in incidence, and reduction in risky behaviour, started as early as 1994.

This is not to diminish the importance of death in the changing face of the HIV epidemic. HIV incidence will decline as a natural direct consequence of the early attrition of that class of society exhibiting the riskiest sexual behaviour. If there is no accompanying change in human behaviour, however, the incidence soon levels off and equilibrium is established when the death rate matches the rate of occurrence of new infections, and the HIV prevalence is then maintained at a high level.

But death can also have an indirect effect on HIV prevalence, if human beings react to the manifest effects of the HIV epidemic by changing their behaviour. And the evidence of massively increased death rates in the recent past is all too obvious in Harare. An area of 150 hectares, just south of Harare, that was bush just ten years ago, is now a massive cemetery, with tens of thousands of closely-packed graves.

Mathematical modelling shows that it is possible to produce a good fit to the HIV prevalence, in Fig. 4, under the assumption that the rate of occurrence of new infections decreases as the observed death rate increases. The best fit of the model to the HIV incidence, prevalence and mortality data suggests, moreover, that the death rate should have peaked in about 2001/2, as indeed appears to be the case from the records of deaths in Harare. The fit is thus consistent with the feasible contention that people have modified their behaviour according to the death rate at any given time. The particular mathematical assumption that incidence declines exponentially with death rate does, however, come with the consequence that reductions

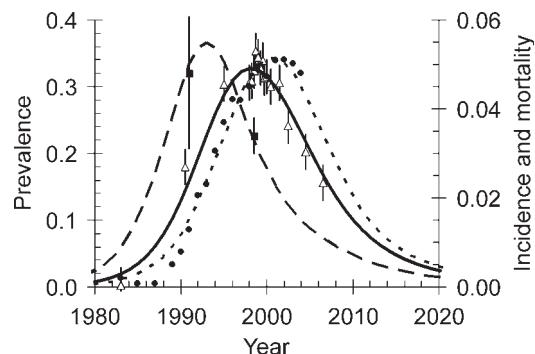


Fig. 4. Estimated (dots, squares and triangles) and fitted (curves) HIV incidence (—), prevalence (—), and deaths (---) among women attending antenatal and maternity clinic in Harare, Zimbabwe. Data from ref. 18.

in risky behaviour lead to declines in the death rate—and thus, by our model, may lead to eventual increases in risky sexual behaviour. The predicted incidence, prevalence and death rates thus level off, instead of continuing to decline steadily (Fig. 4). The model does not, in other words, attribute any 'memory' or 'learning behaviour' to the individuals modelled. Levels of risky behaviour simply increase or decrease mechanically as a simple inverse function of the death rate.

It is to be hoped—and this is my expectation—that this is an unduly misanthropic view of humankind, and that people will actually be (even) more thoughtful in their response to the observed effects of HIV than suggested by our model. Conversely, many would view my perception as an unrealistically optimistic view of the world. Indeed, even the suggestion that HIV incidence has been declining in Harare, and that this must be due to changes in behaviour, has met with great scepticism from some scientists. One is left with the feeling that they find it hard to believe that people are capable of altering their sexual behaviour—except perhaps through enlightened government policy, or via the insightful intervention by a generously funded international organization.

Judith Todd is more optimistic about human behaviour.¹⁷ She notes that Zimbabweans overwhelmingly rejected Ian Smith's offer of a political settlement in 1972, brokered by the British government via the Pearce Commission. Similarly in 2000, they overwhelmingly rejected Robert Mugabe's offer of a new constitution, which brought promise of free farm land. The leadership in both cases drew the conclusion that the electorate was politically naïve and easily led astray by the unscrupulous; black nationalists in the first instance, white farmers in the second. Todd, who has been cruelly persecuted under both Smith's and Mugabe's governments, drew a quite different conclusion. She argues that, in both cases, the people accurately detected evil intent, and rejected it.

Perhaps that same basic human intelligence, rather than the efforts of governments and aid agencies, is the most important factor underlying the decline in HIV prevalence seen not only in Harare, but also in Uganda and increasingly in other countries in the region.¹⁹ Perhaps governments and aid agencies find this pill so hard to swallow, because it devalues their contribution. Nonetheless, if true, what the message implies is that the most important role of the scientist remains exactly as it should be—the objective and accurate observation, analysis, and dissemination of information.

The developing HIV epidemic in South Africa

This more optimistic view might appear at first sight to find little support from the observed trend of the HIV epidemic in South Africa—where, by 2006, there was still no evidence of the beginnings of the rapid decline in HIV prevalence that was observed in Harare (Fig. 5), despite the abundant and direct evidence everywhere in South Africa of the effects of HIV infection. It is true that HIV prevalence for 2006 was marginally (though not significantly) lower than for 2005 and estimates from the ASSA model suggest that HIV incidence has started to decline.²⁰ It is also true (Fig. 5) that the epidemic in South Africa lags the Zimbabwean epidemic by about eight years, so that it is possible (though not in the least guaranteed) that we might be about to witness a similarly sharp decline in HIV prevalence in South Africa.

The present levels of HIV infection everywhere in southern Africa are nonetheless entirely unacceptable. Even in Zimbabwe, where HIV prevalence appears to have been declining rapidly for at least six years, the last published estimates of HIV

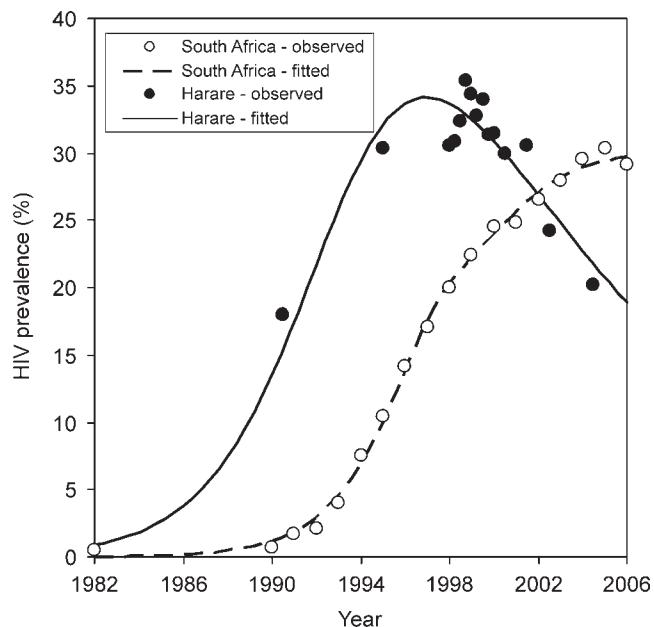


Fig. 5. HIV prevalence among women attending antenatal and maternity clinics in Harare, Zimbabwe,¹⁸ and South Africa (UNAIDS Epidemiology Fact Sheets, 2006).

prevalence in Harare ANCs was still about 20% (Fig. 4)—hardly a healthy situation. Much the same can be said of other southern African countries, and we need to ask what, if anything, we can do to influence this situation in a positive manner.

One small snip for man

There is evidence in some instances of HIV prevalence and incidence declining as a consequence of behavioural change. This has sometimes been attributed to direct governmental action—as was argued when the Thai government pressured female sex workers to use condoms with their clients.²¹ Similarly, declining HIV prevalence in Uganda during the 1990s has been attributed to 'population mobilization', which involved the efforts of central government and civil society.²² The importance of external intervention, other than education regarding the nature and dangers of HIV, is unclear. The declines in HIV incidence and prevalence in Zimbabwe, for instance (Fig. 4), have been of a similar order to those observed in Uganda—but have not been linked to intervention on the scale attributed to the governments of Uganda or Thailand. This area continues to be a matter of debate.

What is not a matter of debate is that there is currently little by way of alternative options in the drive to reduce HIV incidence. Whereas very large sums of money continue to be invested in the development of HIV vaccines, and of microbicides that could be used to block infection biochemically, the sad truth is that there is no pharmaceutical product of any kind that is currently effective in reducing HIV incidence. Organizations centrally involved in the production of such medication are of the opinion that we are years away from delivering even a partially effective vaccine or microbicide.²³

In the absence of viable pharmaceutical protection, the recent publication of the results of three randomized control trials demonstrating a more-than-50% protective effect of male circumcision on female-to-male transmission of HIV (see references above) is important. Mathematical modelling suggests that a full-scale roll-out of male circumcision in sub-Saharan Africa could avert 2.0 (95% CI 1–3.8) million new HIV infections and 0.3 (95% CI 0.2–0.5) million deaths over the next ten years. A further 3.7 (95% CI 1.9–7.5) million new HIV

infections and 2.7 (95% CI 1.5–5.3) million deaths could similarly be averted, over the ten years thereafter.

This modelling involved contributions from several of my associates at SACEMA. The centre has from its inception taken an active interest in the development of male circumcision as an intervention, and continues to support Bertran Auvert, for example, in his development of both the theoretical and practical aspects of the work.

The success of the intervention, as modelled above, depends of course on the assumption that the natural reduction in the probability of female-to-male transmission of HIV in circumcised men is not offset by increases in risky behaviour, tending to reduce, nullify or even over-compensate for the protective effect. The only way to find out is to move towards a more general roll-out of male circumcision as an intervention method and, crucially, to keep track of the circumcised men to compare their HIV incidence relative to men in the general population.

SACEMA fully supports the idea of expanding the application of medical male circumcision in South Africa, as a natural follow-up to randomized control trials. Such operational research will allow further assessment of the challenges associated with applying the procedure on larger scales, while providing the necessary time to assess the longer-term protective effects.

Further evidence of the protective effects of medical male circumcision would inevitably influence cultural leaders of non-circumcising ethnic groups to reassess their attitudes to the practice. Moreover, even among groups that do habitually circumcise their males, it may lead to a reassessment of the timing and means of male circumcision. It has been noted that the absence of clear differences in HIV prevalence rates between South African ethnic groups who do, and do not, circumcise their males may be related to the manner in which the circumcision is carried out. Medical observers note that traditional male circumcision can remove variable amounts of the foreskin and that, therefore, the protective effect of traditional male circumcision against HIV infection may be less than optimal.

If the observed ongoing results of increased frequency of medical male circumcision are sufficiently encouraging, the medical fraternity may need to start asking questions about whether they should be advising parents of all ethnic groups to have their sons medically circumcised at birth as a matter of course. To the objection that this flies in the face of long-standing cultural values, one may counter quite simply that cultures do, and should, change: and there is no better reason to change than in the interests of survival. Regardless of the changes which do, or do not, occur with regard to future male circumcision practices, what is of paramount importance is that all male circumcisions should be carried out with the utmost regard to safety for the patient.

Antiretroviral therapy as an offensive weapon

The above discussion of methods available for reducing HIV incidence and prevalence has centred on prevention of infection. With double-digit HIV prevalence levels throughout southern Africa, however, we have massive numbers of people who are already infected with the virus. There has, quite rightly, been increasing pressure over the years to improve access to antiretroviral therapy—with the main emphasis on the improved survival and quality of life of the patient, consistent with considerations of delaying treatment so as to lessen the time spent on drugs that can have unpleasant side effects and simultaneously reducing costs and the chances of developing resistance. These arguments are used to justify the standard practice in Africa of

starting antiretroviral therapy (ART) only when the CD4 count drops to $200/\mu\text{l}$,²⁵ or when an HIV-positive patient presents with symptoms typical of a WHO stage 3 or 4 infection.

There are several problems with this approach:

- i) As a patient's CD4 count declines, the probability of acquiring opportunistic infections increases. In particular, by the time the count approaches $200/\mu\text{l}$, many will already have tuberculosis, which has to be treated before ART can commence.
- ii) Even if the patient has not acquired TB, if the CD4 count is below $200/\mu\text{l}$ at the onset of ART, this count does not fully recover in some patients. While the *mean* CD4 count increases, the increase in the *mode* is less impressive, and a proportion of patients have CD4 counts that remain below $200/\mu\text{l}$.
- iii) Even after three years on ART, the probability of infection with TB is several times higher than in HIV-negative people.
- iv) These results are consistent with the view that prolonged infection with HIV leads to accelerated senescence of the immune system.²⁴

While none of these lines of evidence may be compelling alone, collectively they do constitute an argument in favour of a markedly earlier administration of antiretroviral therapy—in the interests of the individual HIV-positive patient. And there is a further important epidemiological reason for wanting to start earlier therapy. Analysis of data from one study in the Western Cape province has shown that whereas the roll-out was reaching seven times the national average, in terms of the proportion of patients on ART among those who qualify, the HIV-positive patients who were not receiving ART were predominantly under 30 years of age. This is entirely natural, as the youngest people tend, by definition, to have been HIV positive for the shortest time, and fewer have CD4 counts below $200/\mu\text{l}$ than in older age groups. On the other hand, the young are the most sexually active age group, and thus the most important target for any programme that aims to reduce transmission rates. Earlier onset of ART would have the important and natural epidemiological effect of reducing the viral load in larger numbers of sexually active patients and would thereby automatically reduce the probability of this group infecting their sexual partners. It seems, therefore, that there is a convergence of the individual and population interests in the early recruitment of HIV-positive patients into ART programmes.

If modelling indicates that the aggressive use of ART in this way could be of service in reducing HIV incidence, there may be a further argument for a more active approach towards the identification of HIV-positive cases, such that major reductions in viral load can be effected in the greatest proportion of HIV-positive people. While political considerations currently prevent general compulsory HIV testing, it is interesting to note that Botswana has already adopted an 'opt out' approach to HIV testing. That is to say, anybody visiting a clinic can automatically be given an HIV test—unless they specifically ask not to be tested.

Conclusion

I have discussed various possible measures aimed at addressing the immediate problem of reducing HIV incidence in South Africa: earlier roll-out of antiretroviral therapy, coupled with a more aggressive approach to case detection and promotion, and proactive promotion of male circumcision. Any such intervention will have to be combined with education programmes explaining the reasons for male circumcision and especially the fact that it gives partial but not complete protection, so that the use of

condoms is still vigorously promoted. The roll-out of ART will have to be combined with education and counselling, especially in regard to protecting one's partner and future children, but may also serve as a way of encouraging people to come forward for testing. These interventions will need very substantial commitment and financial resources, and much of the responsibility for this will fall on the shoulders of government. There will undoubtedly be debate about the wisdom of these approaches but, even if they are as successful as we hope they will be, one needs to be mindful of the fact that they essentially amount to the management of an immediate crisis.

None of these interventions—nor previous efforts to encourage abstinence, fidelity or condom use—addresses the fundamental social problems which have ensured that the HIV epidemic is so severe in southern Africa. Indeed, AIDS in southern Africa should be seen not only as a disease but also as a symptom of social ills, in general, and of the fragmentation and breakdown of family life in particular. If the analysis presented here has any validity at all, what it implies is that until we manage to address the underlying problem of the breakdown of family structures in the region—and in South Africa in particular—we must expect that our populations will continue to be prey to HIV, and to any new sexually transmitted infection that arises.

Nor is the problem restricted simply to sexually transmitted infections. South Africa currently has one of the highest rates of violent crime in the world. There are, annually, twice as many murders in Cape Town as there are in the whole of England and Wales. One can point a finger at inadequate policing, corruption and a general lack of political will in combating crime. But it is tempting to suggest that the high crime levels and the high incidence of HIV infection have the same structural source—the breakdown of family structure associated with oscillatory migration.

In combating HIV/AIDS there is, of course, a continuing imperative for scientific innovation, for education and mobilization involving religious, community and political leadership, and for increased efforts in providing treatment for those already infected. But the underlying cause of the HIV epidemic, and of many of our social ills, has its roots in policies set in motion more than a century ago. The present South African government is not responsible for those policies. It is, however, in the unenviable position of having inherited the responsibility for the consequences—and therefore needs to mobilize all appropriate help in addressing the underlying cause of the HIV epidemic in this country.

I am very grateful to Simon Bekker and David Sanders for discussions on matters relating to this article, to Ekkehard Kopp, Margaret Ward and Edwin Hees for correcting and improving the manuscript, and to Mattie van der Merwe for additional help. I thank Eleanor Gouws for supplying me with data. Brian Williams has been hugely helpful throughout the production of this piece and conceived many of the ideas presented here. Any errors of fact and interpretation, however, remain entirely my own.

1. Drain P.K., Smith J.S., Hughes J.P., Halperin D.T. & Holmes, K.K. (2004). Correlates of national HIV seroprevalence. An ecologic analysis of 122 developing countries. *J. Acquir. Immune Defic. Syndr.* **35**, 407–420.
2. Drain P.K., Halperin D.T., Hughes J.P., Klausner J.D. & Bailey R.C. (2006). Male circumcision, religion and infectious diseases: an ecologic analysis of 118 developing countries. *BMC Infectious Diseases* **6**, 172. doi:10.1186/1471-2334-6-172. Online at: www.biomedcentral.com/1471-2334/6/172
3. Werker E., Ahuja A. and Wendell B. (2006). Male circumcision and AIDS: the macroeconomic impact of a health crisis. *Harvard Business School Working Paper* No. 07-025.
4. Piot P., Greener R. and Russell S. (2007). Squaring the circle: AIDS, poverty, and human development. *PLoS Med* **4**(10): e314. doi:10.1371/journal.pmed.0040314
5. UNAIDS (2006). Epidemiology Fact Sheets. UNAIDS, Geneva.
6. Williams B.G., Lloyd-Smith J.O., Gouws E., Hankins C., Getz W.M. *et al.* (2006). The potential impact of male circumcision on HIV in sub-Saharan African populations. *PLoS Medicine* **3**(7): e262. doi: 10.1371/journal.pmed.0030262
7. Auvert B., Taljaard D., Lagarde E., Sobngwi-Tambekou J. *et al.* (2005). Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 trial. *PLoS Medicine* **2**: e298. doi:10.1371/journal.pmed.0020298
8. Gray R.H., Kigozi G., Serwadda D., Mukaumbi F. *et al.* (2007). Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. *Lancet* **369**, 657–666.
9. Bailey R.C., Moses S., Parker C.B., Kawango A. *et al.* (2007). Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised control trial. *Lancet* **369**, 643–656.
10. Hargrove J. and Williams B. (2007). Response to Piot *et al.* (2007). Squaring the circle: AIDS, poverty, and human development. *PLoS Med* **4**(10): e314. doi:10.1371/journal.pmed.0040314
11. Sanders D. and Sambo A. (1991). AIDS in Africa: the implications of economic recession and structural adjustment. *Health Policy and Planning* **6**, 157–165.
12. Lurie M. (2000). Migration and AIDS in Southern Africa: A review (2000). *S. Afr. J. Sci.* **96**, 343–347.
13. Carswell J.W., Lloyd G. and Howells J. (1989). Prevalence of HIV-1 in east African lorry drivers. *AIDS* **11**, 759–761.
14. Pakenham T. (1979). *The Boer War*. Weidenfeld and Nicholson, London.
15. Epstein H. (2007). *The Invisible Cure: Africa, the West, and the Fight against AIDS*. Penguin, London.
16. Humphrey J.H., Nathoo K.J., Hargrove J.W., Iliff P.J. *et al.* (2007). HIV-1 and HIV-2 prevalence and associated risk factors among postnatal women in Harare, Zimbabwe. *Int. J. Epidemiol.* **135**, 933–942.
17. Todd J.G. (2007). *Through the Darkness: A Life in Zimbabwe*. Struik, Cape Town.
18. Mahomva A.I., Greby S., Dube S. *et al.* (2006). HIV prevalence and trends from data in Zimbabwe, 1997–2004. *Sexually Transmitted Infections* **82**, i42–i47.
19. Hallett T.B., Aberle-Grasse J., Bello G. *et al.* (2006). Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. *Sexually Transmitted Infections* **82**, 1–8.
20. Dorrington R.E., Johnson L.F., Bradshaw D. and Daniel T. (2006). *The Demographic Impact of HIV/AIDS in South Africa. National and Provincial Indicators for 2006*. Centre for Actuarial Research, South African Medical Research Council and Actuarial Society of South Africa, Cape Town.
21. Nelson K.E., Celentano D.D., Eiumtrakol S., Hoover D.R. *et al.* (1996). Changes in sexual behavior and a decline in HIV infection among young men in Thailand. *N. Engl. J. Med.* **335**, 297–303.
22. Stoneburner R.L. and Low-Beer D. (2004). Population-level HIV declines and behavioral risk avoidance in Uganda. *Science* **304**, 714–718.
23. AVAC (2007). Aids Vaccine Advocacy Coalition Report (2007). *Resetting the clock*. Online: www.avac.org
24. Appay V., Almeida J.R., Sauce D. *et al.* (2007). Accelerated immune senescence and HIV-1 infection. *Exp. Geront.* **42**, 432–437.
25. CD4 cells (or T cells) are lymphocytes that act as receptors of HIV in the immune system. In a healthy adult a normal CD4 count is typically 600–1200 cells per ml of blood. HIV infection reduces the count. For a count of 200, the immune system is so severely weakened by loss of CD4 cells that an HIV-positive person is at much enhanced risk of opportunistic infection.

This article is accompanied by supplementary material at www.sacema.ac.za