SAMJ
Articles

Understanding the epidemic of HIV in South Africa

Analysis of the antenatal clinic survey data

Brian Williams, Catherine Campbell

Objectives. To investigate the magnitude and the time course of the HIV epidemic in the provinces of South Africa from the antenatal clinic HIV surveys.

Design. We analysed the data on the provincial prevalences of HIV infection from 1990 to 1996 using maximum likelihood methods to determine the intrinsic growth rate and probable asymptotic prevalence of HIV among women attending antenatal clinics.

Subjects. Women attending antenatal clinics and included in the national HIV prevalence surveys conducted by the Department of Health.

Results. 1. In KwaZulu-Natal the epidemic is likely to peak at a prevalence of about 23% (95% confidence interval (CI) 19 - 36%). 2. The intrinsic doubling time does not differ significantly among the provinces. 3. The average length of the intrinsic doubling time is 12.0 months (95% CI 11.3 - 12.8 months). 4. The force of infection is approximately 1.00/year at age 16 years and declines at a rate of about 5% per year of age above 16 years.

Conclusions. South Africa is likely to experience one of the worst HIV epidemics in Africa. The lack of statistically significant differences between the growth rates of the epidemic in the various provinces constrains the possible explanations that can be advanced to explain the time course of the epidemic and may in part be a consequence of migrancy. The intrinsic growth rate is higher than previous estimates and it is possible that in those provinces where the prevalence is still low it will eventually reach the same levels as in KwaZulu-Natal.

S Afr Med J 1998; **88:** 247-251.

The data collected by the Department of Health through their annual antenatal clinic surveys represent the most important source of information on HIV in South Africa. These are the only data that provide any indication of the time course and geographical spread of the epidemic in this country.[1,2] The importance of these data cannot be overstated, but it is essential to analyse them carefully if reliable conclusions are to be drawn about the past and probable future time course of the epidemic. We use the data to estimate the likely asymptotic prevalence, the intrinsic growth rate of the epidemic in different provinces and the age-dependent force of infection. The asymptotic prevalence gives a measure of the probable long-term burden of disease. The intrinsic growth rate indicates the rate at which the infection is increasing and provincial estimates will partly determine explanations of the spread of the epidemic. The age-dependent force of infection indicates how the risk of infection varies across age groups. When serious attempts to manage the epidemic are made, these data and the parameters derived will provide essential baseline information against which to assess the effectiveness or otherwise of disease management programmes.

Epidemiology Research Unit, PO Box 30606, Braamfontein, 2017
Brian Williams, PhD (Cantab)
London School of Economics, Houghton Street, London, UK
Catherine Campbell, PhD (Bristol)

The dataset

The data are taken from the results of the national HIV surveys.[1,2] These give, by province, the number of women tested and the proportion who are HIV-positive each year from 1990 to 1996. Two limitations of these data for making predictions are immediately evident. Firstly, for each province there are only seven data points, a feature that limits the precision with which model parameters can be estimated, and secondly in some of the provincial surveys, especially in the earlier surveys, only 1 or 2 people were positive, which results in very large error margins on some of these data points.

The logistic model

The 1996 survey gives, for the first time, evidence of a decline in the growth rate of the epidemic. We have therefore chosen to fit the data to logistic curves from which the intrinsic growth rate of the epidemic (the rate of increase in prevalence when the prevalence is still very low) and of the asymptotic prevalence (the prevalence at which the epidemic will level off) can be estimated. To fit the data we have used maximum likelihood estimation procedures,[3] which give minimum variance estimators for the parameters and, in particular, allow the use of exact binomial error distributions, which are especially important when the numbers are small.

Estimating the asymptotic prevalence

Each of the provincial datasets was fitted to the logistic model given in Appendix 1 and the resulting asymptotic values are shown in Fig. 1. (Details of the calculation of the error limits are given in Appendix 2.) From Fig. 1 it is clear that the epidemic is sufficiently advanced to obtain useful estimates of the asymptotic prevalence only in KwaZulu-Natal and Mpumalanga. For KwaZulu-Natal the best fit value is 23.4% (18.7 - 35.6%, 95% confidence interval (CI) here

and elsewhere). For Mpumalanga the estimated asymptote is 19.6% (16.1 - 26.9%). Apart from Northern Province, for which the data are problematic, as discussed below, there are no significant differences between provinces and we have set the asymptotic prevalence at 26.8% in the following analyses. This is consistent with rates in other countries in the region where, for example, the antenatal prevalence in 1995 was 29% in Gaborone, Botswana, 19% in Swaziland and 32% in Harare, Zimbabwe (US Bureau of Census, Population Division, International Programmes Center, HIV/AIDS Surveillance Database, June 1996).

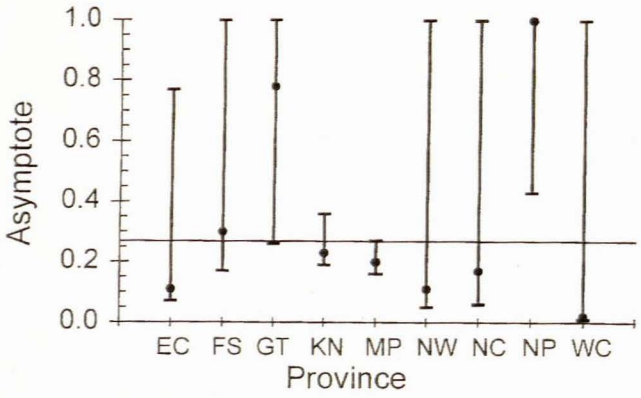

Fig. 1. The asymptotic prevalence of HIV infection in the various provinces. EC: Eastern Cape, 11% (7% - 77%); FS: Free State, 30% (17% - 100%); GT: Gauteng, 78% (26% - 100%); KN: KwaZulu-Natal, 23% (19% - 26%); MP: Mpumalanga, 20% (16% - 27%); NW: North-West, 100% (> 43%); NC: Northern Cape, 11% (5% - 100%); NP: Northern Province, 17% (6% - 100%); WC: Western Cape, 2.4% (1.3% - 100%). The error bars give 95% confidence limits. The horizontal line is the value used in subsequent calculations.

Estimating the intrinsic rate of increase (doubling time)

It has become something of a received truth in South African HIV/AIDS circles that the epidemic is progressing faster in some provinces, e.g. KwaZulu-Natal, than in others, e.g. Western Cape. Do the data support this assumption? Having fixed the asymptotic prevalence to 26.8% in all provinces, logistic curves were then re-fitted to each provincial dataset with the results shown in Fig. 2.

With the exception of Mpumalanga and North-West, the fits are acceptably good in all cases. In Mpumalanga the data seem unreliable as they imply that there was no significant increase in prevalence from 1992 to 1993 followed by a fivefold increase in 1994. In North-West, the prevalence for 1991 would appear to be in error and while the increase from 1994 to 1995 was marginal and not statistically significant, there appears to have been an almost threefold increase in 1996. It would seem likely that the 1995 value is incorrect.

From the fits given in Fig. 2, the intrinsic rate of increase can be estimated (r in equation 1, Appendix 1) and the initial doubling times for the epidemic in each province can be estimated as shown in Fig. 3. None of the doubling times differs significantly from the mean value, which is 12.0 months (11.3 - 12.8 months). The fact that there are no significant differences among the doubling times in the different provinces is surprising, as one might expect the epidemic pattern to vary between places with different socio-economic conditions. In the Western Cape, however, where the prevalence is low and few people in the sample have tested positive, the error bars on the data (Fig. 2) are correspondingly large and the estimate of the doubling time correspondingly poor; future data might well show that the epidemic is increasing more slowly in the Western Cape than elsewhere. On the basis of the available data the most parsimonious assumption is that the epidemics are following essentially the same time course but lagging behind KwaZulu-Natal.

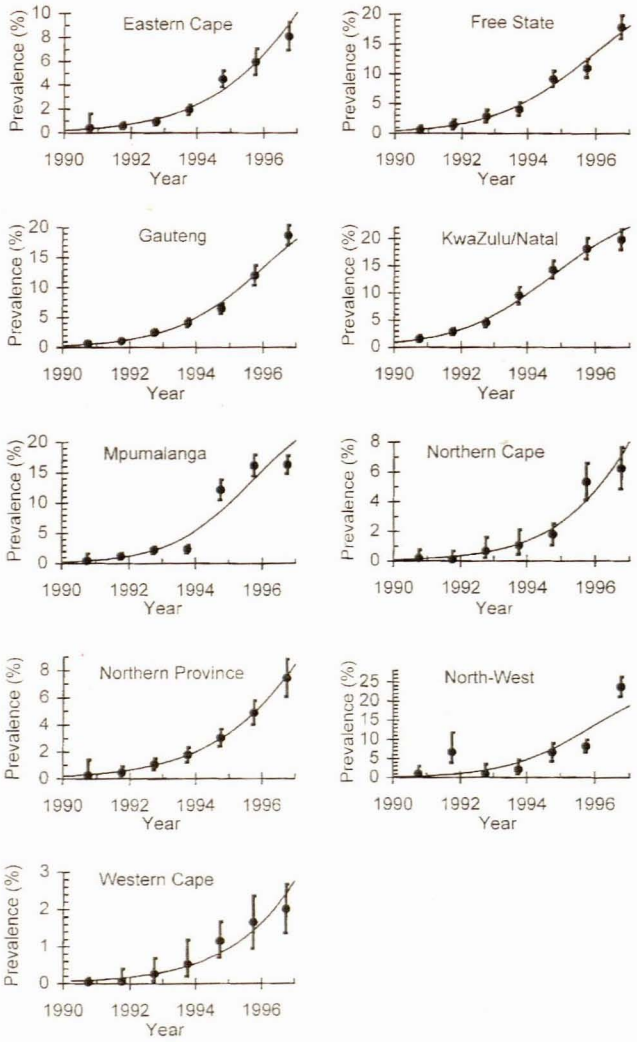

Fig. 2. Logistic curves fitted to the provincial antenatal clinic data with the asymptotes set to 26.8%.

In order to determine the time lag between the epidemics in the various provinces we set the intrinsic growth rate for each province to the average value for all the provinces (0.691 ± 0.066 per year), keeping the asymptotic prevalence at the previous value of 26.8%; we then fitted each provincial dataset to logistic curves. The results in Fig. 4 show that Mpumalanga, Free State, North-West and Gauteng all follow the epidemic in KwaZulu-Natal by about 1

year, the Eastern Cape, Northern Province and the Northern Cape by about 3 years and the Western Cape by about 5 years.

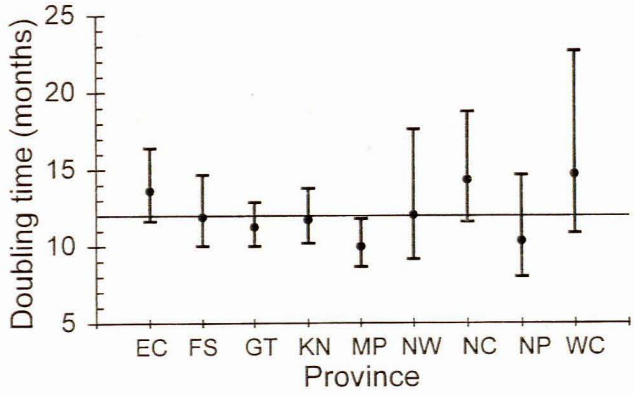

Fig. 3. Doubling times for the prevalence of HIV based on the intrinsic rate of increase in the various provinces obtained by fitting logistic curves to the provincial prevalences with the asymptote set to 26.8%. EC: 13.6 months (11.7 - 16.4 months); FS: 11.9 months (10.0 - 14.7 months); GT: 11.3 months (10.0 - 12.9 months); KN: 11.7 months (10.2 - 13.8 months); MP: 10.0 months (8.7 - 11.8 months); NW: 12.1 months (9.2 - 17.6 months); NC: 14.3 months (11.6 - 18.8 months); NP: 10.4 months (8.0 - 14.7 months); WC: 14.7 months (10.9 - 22.7 months).

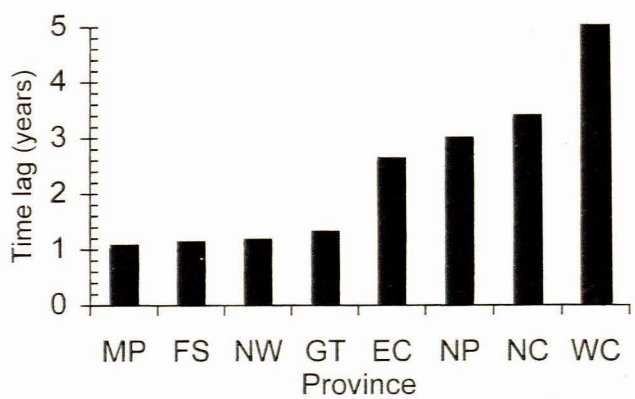

Fig. 4. The time lag in years between the epidemics in the various provinces if it is assumed that the asymptotic value is 26.8% and the intrinsic growth rate is 0.691 per year in all cases.

Estimating the force of infection

Using the age distribution of prevalence, it is possible to estimate the force of infection (λ) which, when multiplied by the prevalence, gives the rate at which those susceptible become infected. The model, given in Appendix 3, assumes that the force of infection is proportional to the overall prevalence at any time but declines exponentially with age. Unfortunately the numbers of people in each age group who were tested in each province are too small to estimate the force of infection for each province separately, and we have therefore fitted the model to the national age prevalence data for 1995 and 1996 (Fig. 5).

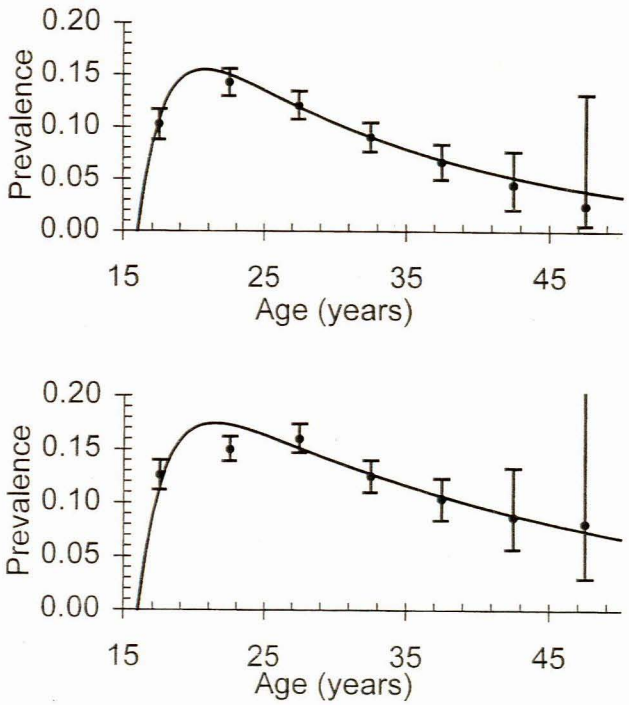

Fig. 5. The national age prevalence for HIV infection from the antenatal clinic data fitted to the model given in Appendix 3 (upper curve 1995, lower curve 1996).

The best fits to the data are obtained if we assume that sexual activity first takes place at 16 years of age. The estimates of the force of infection are then 1.05 ± 0.20 per year for 1995 and 0.91 ± 0.20 per year for 1996 at age 16, and the estimates of the rate at which the force of infection decreases with age are 5.8 ± 1.9% per year for 1995 and 3.8 ± 1.9% per year for 1996. In fact it is likely that some people engage in sexual activity before the age of 16 and that the risk of infection increases with age to some maximum value, after which it declines. The present data do not warrant the use of more sophisticated models. The important point is that if sexual activity does not decline with age, the decline in prevalence with age cannot be reproduced; however, the rate at which prevalence declines with age also depends on the rate at which the epidemic increases. In essence the prevalence among people who are now aged 50, say, is low both because the likelihood of their now engaging in high-risk sexual activities is less than it is for the younger age groups, and because when they were younger the prevalence was lower, which consequently lowered their risk. Based on the available data, at the end of 1996 a 16-year-old in KwaZulu-Natal had an annual risk of becoming infected of about 20%, while for a 46-year-old the annual risk was about 4%. In the Western Cape the corresponding risks were about 2% and 0.4%.

Discussion

This analysis of the provincial data for the prevalence of HIV infection among women attending antenatal clinics indicates that the best estimate of the eventual asymptotic prevalence

in KwaZulu-Natal is 24%. Reasonably precise estimates of the prevalence can only be obtained for KwaZulu-Natal, because the data for North-West and Mpumalanga are problematic and the epidemics in the other provinces are not sufficiently advanced. However, the data for all of the provinces are consistent with an asymptotic prevalence of 27% and this is similar to values that have already been observed in other cities and countries in southern Africa (US Bureau of Census, Population Division, International Programmes Center, HIV/AIDS Surveillance Database, June 1996). It should be noted, however, that more detailed models[4] generally show that the prevalence initially increases exponentially, peaks as the number of remaining susceptible individuals declines, and then falls by up to 10 - 20% as those with HIV infection succumb to AIDS and AIDS-related diseases. We do not yet have sufficient data to test these more sophisticated models rigorously. If the nationwide prevalence among women attending antenatal clinics does reach this level, the number of HIV-infected people could reach approximately 4.5 million, which implies that as the epidemic matures there could be between 500 000 and 1 million deaths a year from AIDS and AIDS-related disease. This will place a substantial burden on caregivers and the health services generally and plans must be implemented as a matter of urgency to deal with this. Furthermore, relatively few data are available on the burden and clinical spectrum of AIDS-related disease among various groups of people in South Africa, without which it is almost impossible for the health services to plan ways in which to deal with the consequences of the epidemic.

The lack of significant differences among the intrinsic epidemic growth rates in the various provinces is surprising, although the data do not rule out the possibility that more extensive data collected in future surveys will reveal such differences. The very high levels of migrancy in South Africa must be important in determining the patterns of disease transmission, but these have received very little attention. A possible explanation for the similarity between the doubling times in the various provinces is that the effect of migrancy between provinces is to lock the epidemics together, so that the epidemic in the low-prevalence areas is effectively driven by the epidemic in the high-prevalence areas. Consider, for example, two areas, in the first of which the doubling time is 1 year and in the second of which it is 2 years. If it is assumed that 10% of the people in each area move to the other area each year the epidemic growth rate in the first area falls slightly and in the second area increases rapidly to match that in the first area.[5] Much more extensive models of the geographical distribution and spread of HIV are needed to elucidate this, and the antenatal clinic data which are now available for each health district in the country will provide an important starting point for such analyses.

The estimates of the force of infection are still speculative, given that the extent of the data is insufficient to provide precise estimates. However, the available estimates and, in particular, the way in which they vary with age, provide an important starting point for modelling the impact that the epidemic is likely to have on different age groups and this will be essential both for deciding on how best to target intervention programmes and for reliable estimation of the likely social and economic impacts of the epidemic.

Having done little to stem the rising tide of the epidemic over the last 7 years, we now read that 'with HIV infection already so high . . . some activists say "we missed the boat" on prevention and that resources should now be focused on providing the best possible care'.[6] The increase in the care needs of people with AIDS and AIDS-related diseases will undoubtedly lead the health services to change the way in which resources are allocated and patients are dealt with. But it is important to stress that even though the prevalence in KwaZulu-Natal and some other provinces may be approaching their peak levels, currently uninfected people will still become infected while some of those who are already infected will die, and a reduction in the force of infection will still prevent many future deaths.

While the antenatal clinic data are of great importance in providing information about the present state and probable future spread of the epidemic, it is remarkable that they are still virtually the only useful data available. Resources should be committed to the supplementation of these data with much more detailed studies at sentinel sites, which should be chosen to be representative of different socio-economic situations in the country. These sentinel sites should be used to monitor not only HIV but other sexually transmitted diseases that enhance the spread of HIV, and they should be used to collect socio-economic and behavioural data, especially those on sexual networking patterns; this information should be linked to the biomedical data. Such data are now being collected in Hlabisa. Furthermore, a group representing a range of stakeholders including three mining houses, the trade unions and the Department of Health, as well as research institutes and non-government organisations, is planning a major intervention to improve the management of STDs and to promote community outreach and condom distribution in Carletonville. This intervention will be carefully monitored and evaluated using biomedical measures of STD incidence and HIV prevalence as well as social measures of behaviour, knowledge and attitudes towards HIV and STDs. Biomedical and behavioural surveys will be carried out at the beginning and at the end of the project and, during the course of the project, extensive in-depth interviews and focus group discussions will be held with those living and working on the mines as well as the people in the Carletonville community. It is hoped that such studies will supplement the few data that are currently available and provide an important contribution to the management of HIV and STDs in other settings.

Appendix 1. With a logistic model, $p(t)$, the prevalence of infection at time t may be written

$$p = \frac{p_\infty p_0 e^{r(t - t_0)}}{(p_\infty - p_0) + p_0 e^{r(t - t_0)}}$$

where p_∞ is the eventual asymptotic prevalence, p_0 is the prevalence at time t_0, and r is the intrinsic growth rate. In the early stages the prevalence increases exponentially at a rate r and then converges exponentially to p_∞, also at a rate r.

Appendix 2. When the error limits on the estimates are calculated, the three parameters p_∞, p_0 and r are closely correlated. Furthermore, p_0, in particular, is bound from below at 0, while the coefficient of variation is, in many cases, close to 1. The standard error analysis using

maximum likelihood techniques assumes that the likelihood function is Gaussian while in these analyses the likelihood function, and the consequent error limits, are highly asymmetrical. In order to estimate the upper and lower confidence limits on the asymptotic prevalence it is therefore necessary to vary the asymptote, maximise the likelihood function by varying the other two parameters, p_0 and r, and use a likelihood ratio test to find critical values corresponding to upper and lower bounds for 95% confidence limits.

Appendix 3. In order to obtain a functional form for the age-prevalence data, we note that the probability that a susceptible individual becomes infected at any time is proportional to the force of infection, λ, multiplied by the overall population prevalence. Consider then a person whose age is a_1 at time t_1. The probability that this person would have been infected at age a (assuming they were not already infected) is

$$p(a) = \lambda P\,(t_1 - a_1 + a) = \lambda c e^{r(t_1 - a_1 + a)}$$

where $P(t)$ is the prevalence at time t, and c and r are chosen to match the change in overall prevalence with time. If we also let sexual activity decline exponentially from 1 at age a_0 at a rate s, equation 2 becomes

$$p(a) = \lambda P\,(t - a_1 - a) e^{-s(a - a_0)}.$$

Then, counting in time increments of δt, the probability of not being infected between ages a_0 and a_1, is:

$$q = (1 - \lambda c e^{r\tilde{t}}\,\delta t)(1 - \lambda c e^{r\tilde{t}} e^{(r-s)\delta t}\,\delta t)...(1 - \lambda c e^{r\tilde{t}} e^{(r-s)(a_1 - a_0)}\delta t)$$

where $\tilde{t} = t - (a_1 - a_0)$, so that

$$\ln(q) = -\lambda c e^{r\tilde{t}}\,\delta t(1 + e^{(r-s)\delta t} + ... + e^{(r-s)(a_1 - a_0)}).$$

Summing the series gives

$$\ln(q) = \lambda c e^{r\tilde{t}}\,(1 - e^{(r-s)(a_1 - a_0)})/(r - s)$$

and the probability that an individual of age a is infected at time t is $p = 1 - q$.

More sophisticated models would use a logistic form for the time dependence of the prevalence and include the distribution of the relative ages of sexual partners. However, the present data fit this simple model reasonably well and do not warrant the use of more extensive models.

We thank Ronelle Swanevelder for her comments on the manuscript and would especially like to acknowledge the vital role that she played in supervising and analysing the antenatal clinic survey data.

REFERENCES

1. Sixth national HIV survey of women attending clinics of the public health services in the Republic of South Africa, October 1996. *Epidemiological Comments* 1996; **23:** 3-17.
2. Seventh national HIV survey of women attending antenatal clinics of the public health services, October/November 1996. *Epidemiological Comments* (in press).
3. Williams BG, Dye C. Maximum likelihood for parasitologists. *Parasitology Today* 1994; **10:** 489-493.
4. Anderson RM, May RM. *Infectious Diseases of Humans*. Oxford: Oxford University Press, 1992.
5. Williams BH, Campbell CM. Is the HIV epidemic in South Africa reaching saturation? Paper presented at the 18th African Health Sciences Congress in collaboration with the 14th Epidemiology Society of Southern Africa Conference, Cape Town, 14 - 18 April 1997.
6. Day J. Bureaucrats dither as HIV invades. *Mail and Guardian* 1997; 2-8 May: 6.

Accepted 16 May 1997.