Patterns of infection: using age prevalence data to understand the epidemic of HIV in South Africa

B.G. Williams[a*], E. Gouws[b], M. Colvin[b], F. Sitas[c], G. Ramjee[b] and S.S. Abdool Karim[b]

South Africa is experiencing an explosive epidemic of HIV/AIDS, with about one in four women attending ante-natal clinics nationwide being HIV-positive. In order to understand the natural history of the epidemic, to design and target interventions to manage it and to evaluate the impact of interventions that are implemented, it is essential to gather information on the patterns of infection. In particular it is important to know how these vary with gender, age, migrancy status and between urban and rural settings. Ideally, one should measure age-specific incidence but this is difficult to do. Many datasets are available, however, on age-specific prevalence of infection and these are used to investigate the risk of infection with age among a number of different populations. The populations under consideration include women attending ante-natal clinics, urban and rural populations, migrant workers and commercial sex workers. Data are also presented from one work-based survey and from a study of cancer patients at a major hospital in Soweto.

We were able to identify four different patterns of infection among a) women attending ante-natal clinics; b) women in the general population; c) men in the general population; and d) migrant workers. It is interesting that we were unable to show differences between urban and rural populations. Furthermore, the patterns of infection appear to be fairly constant over time, although as the epidemic saturates and reaches a steady state this must change.These data highlight, in particular, the extremely high risk of infection among 15–25-year-old women and among migrant workers of all ages. They should serve not only to highlight the urgency of the situation and the need to deal with the spread of infection effectively, but should also provide a basis for detailed epidemiological modelling, which can be used to predict the future course of the epidemic, plan an effective response and evaluate the impact of interventions.

The prevalence of HIV infection among women attending ante-natal clinics had reached 24% at the end of 1999, with provincial prevalences ranging from a low of 7% in the Western Cape to a high of 33% in KwaZulu-Natal.[1] At the start of the epidemic, the intrinsic doubling time for infection among women attending ante-natal clinics was between one and two years,[2] although the infection rates now appears to be levelling off.[1] If we are to develop ways to deal with the epidemic effectively, make projections as to the likely future course of the epidemic and evaluate the impact of interventions to manage it, it is essential to understand the spread of the epidemic in space and in time.

In this paper we present a range of data on the age-specific prevalence of infection amongst men and women in urban and rural areas and among migrant men, which reflect the age-specific risk of infection. Ideally, we would like to measure the age-specific incidence of infection directly but this is difficult to do. Cohort studies are time consuming, expensive to carry out and hard to sustain and new methods based on detuned ELISA tests have yet to be proven.[3] Recently, however, Williams *et al.*[4] have shown that, with sufficiently precise data, cross-sectional measurements of age-specific prevalences can be used to estimate age-specific incidence and that in a relatively young but rapidly growing epidemic, the shape of the age-specific prevalence and incidence curves are similar, so that the age-specific prevalence gives a first approximation to the age-specific incidence with the latter being shifted to slightly younger ages than the former.

The datasets that we examine include national ante-natal clinic data for South Africa; ante-natal clinic data for a rural area in KwaZulu-Natal; two community surveys, one rural and the other urban; data on migrant mineworkers and truck drivers; data on commercial sex workers; and a study of cancer patients in a major urban centre. We are able to identify four different patterns of prevalence: among women attending ante-natal clinics, among men and women in the general population, and among migrant workers. We consider the implications for the spread and management of the epidemic in South Africa.

Methods

To compare the different datasets, it is necessary to parameterize the data by fitting the age-prevalence curves to suitable functions. As will be seen, in some cases the age-prevalence curve is (statistically) either constant or varies linearly with age, and in all cases we first fit a linear relationship. In those cases for which a linear fit is not statistically acceptable, the age-prevalence increases rapidly between the ages of 15 and 25 or 30 years and declines more slowly with age thereafter. This is consistent with an onset of sexual activity in the early teenage years and a peak of sexual activity between the ages of about 20 and 30 years, as has been demonstrated in a study carried out in Carletonville[5] for which data are presented below. It is also supported by data on the age-incidence of pregnancy in Hlabisa, a rural district of northern KwaZulu-Natal, the most populous of South Africa's nine provinces. These data, shown in Fig. 1, were collected as part of an HIV-seroprevalence survey of 3163 women, age 15–49 years, attending ante-natal clinics in Hlabisa between January and July 1998. Data from the 1996 census were used as the denominator.[4] In order to parameterize the age-risk function, we therefore seek a function that is zero, or very small, before the age of onset of sexual activity, increases rapidly as sexual activity increases and then decreases more slowly among older people. A function which has these properties is the log-normal

$$R(a) = \frac{N}{\sigma\sqrt{2\pi}(a-a_0)} e^{-[\ln(a-a_0)-m]^2/2\sigma^2}, \qquad (1)$$

with offset a_0, mean m, scale parameter σ and normalized to N. A log-normal function gives a good fit to the age-specific fertility

[a]CSIR Miningtek, P.O. Box 91230, Auckland Park, 2006 South Africa.
E-mail: bgwillia@csir.co.za
[b]Medical Research Council, P.O. Box 17120, Congella, 4013 South Africa.
[c]National Cancer Registry and Cancer Epidemiology Unit, SAIMR, P.O. Box 1038, Johannesburg, 2000 South Africa.
*Author for correspondence.

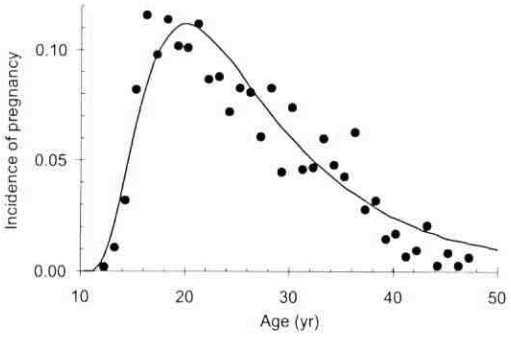

Fig. 1. Fitted (up to 40 years of age) and observed annual fertility estimated from the number of women attending ante-natal clinics in Hlabisa. The curve is log-normal [Equation (1)] with $m = 14.9$ years, $a_0 = 10$ years, $\sigma = 0.629$ and $N = 2.156$. The incidence of pregnancy is shifted down by 0.75 years to allow for the nine months of gestation.

data for Hlabisa as shown in Fig. 1 and, as shown below, gives good fits to all of the datasets for which the age-prevalence of infection deviates from a straight line.

The fitting uses a maximum likelihood procedure,[7] with binomial errors and this makes it possible to obtain the covariance matrices for the fitted parameters from which confidence ellipses can be constructed and used in the comparison of the different fits. These procedures are illustrated here with reference to the national ante-natal clinic surveys.

Table 1 gives the age-prevalence data for each of the ante-natal clinic surveys carried out between 1995 and 1999[1,8] as well as the values of the parameters for the log-normal fits to the data. For the interpretation of these data the mode is more useful than the mean and the fits were done using the mode as one of the parameters. For a log-normal function the relationship between the mode and the mean is

$$m = (\text{mode} - \text{offset}) \times e^{\sigma^2} . \qquad (2)$$

The data and the fitted curves are shown in Figs 2a–e. The P values for goodness of fit are calculated using a χ^2 test of deviance and for all of the data presented in Fig. 2 the significance level, P, is greater than 0.05 (Table 1).

The maximum likelihood fitting procedure can also be used to obtain the covariance matrix of the fitted parameters,[7] from which 95% confidence ellipses for the scale parameter and the mode can be calculated as shown in Fig. 2f. We consider only these two parameters, and not the normalization parameter, since we expect the prevalence of infection to increase with time and we are interested here only in the shape of the age-prevalence curves. Figure 2f shows that the shapes of the curves for 1995, 1997 and 1998 are not significantly different. The data for 1996 are quite different from the rest with a much broader curve (large value of σ) and this can be seen directly by comparing the data in Fig. 2b with the other information in Fig. 2. The data for 1996 were therefore excluded from further consideration. The findings for 1999 show a significant upward shift in the position of the mode, while the shape of the distribution remains the same (Fig. 2f), with the mode now at about 1.5 years of age greater than for data recorded in the previous years.

Data sources

The data used for the analysis in this paper come from a variety of sources and these are described briefly in this section. Further details are provided in the references.

National ante-natal clinic surveys

In October of each year since 1990, the National Department of Health has carried out a prevalence survey among women attending state-run ante-natal clinics throughout the country.[1] Data are recorded by age and province but, unfortunately, the sample size (typically about 15 000 women across the country) is too small to allow for a geographical analysis at less than provincial level. Here we analyse the age prevalence of infection for the entire sample.

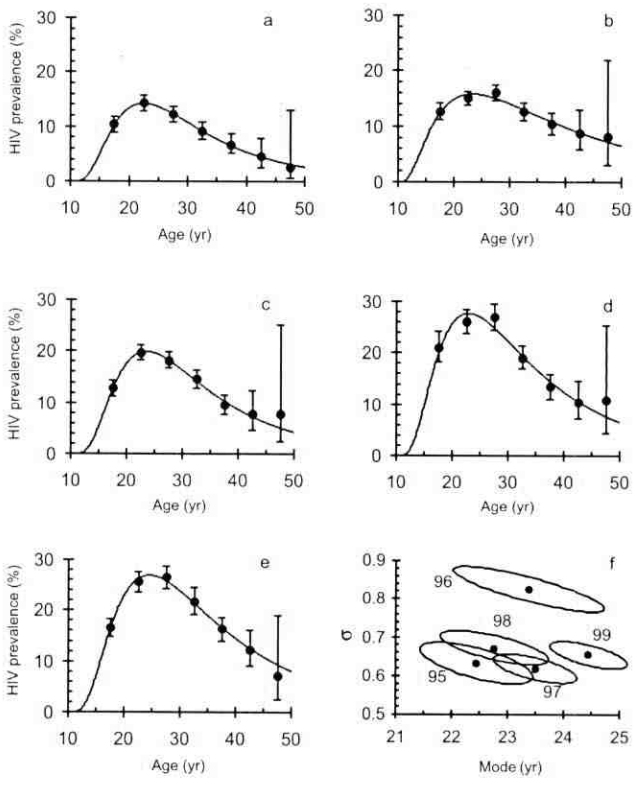

Fig. 2. Age-prevalence of infection for women attending ante-natal clinics nationwide in the annual national surveys fitted to log-normal functions. (**a**) to (**e**) 1995 to 1999, respectively; (**f**) 95% confidence ellipses for the modes and scale parameters of the fits. The data, the parameters of the fits and the goodness of fit are given in Table 1. Error bars are 95% binomial confidence limits.

Ante-natal clinic surveys in Hlabisa

Hlabisa lies about 300 km northeast of Durban. The district is home to about 215 000 predominantly Zulu-speaking people. Since 1992, it has been the site of much of the South African Medical Research Council's HIV and STD research. Hlabisa has a well-developed clinical service, including 15 permanent primary health care clinics, a mobile service and the Hlabisa rural hospital. Anonymous HIV sero-surveys have been conducted among prenatal clinic attendees in Hlabisa since 1992. Further details of the sampling procedures are given in Wilkinson *et al.*[6]

Rural population survey

Three cross-sectional, anonymous, community-based surveys, linked to the malaria surveillance programme, were carried out between 1990 and 1992 in a rural area of northern KwaZulu-Natal. The purpose of these surveys was to determine the prevalence of HIV infection and to monitor the temporal trends and gender differences in the HIV epidemic in rural KwaZulu-Natal. The study population was the same as the target population residing in the area covered by the malaria control programme. Urban areas are excluded from the active malaria surveillance programmes, so the study region was predominantly rural with both subsistence and commercial farms. The survey procedure is described in refs 9 and 10.

Table 1. Age-prevalence of infection among women attending ante-natal clinics in the annual national surveys.[1] The first two columns give the parameters of the log-normal fits and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences. For 1997, 1998 and 1999 the raw data were not available and the numbers were estimated from the given prevalences and confidence limits.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
Year	1995	15–19	1795	184	0.103	0.103
n	3.417	20–24	2828	404	0.143	0.142
Mode	22.447	25–29	2281	277	0.121	0.123
s	0.632	30–34	1574	143	0.091	0.092
m	18.564	35–39	821	55	0.067	0.065
P	0.996	40–44	246	11	0.045	0.045
		45–49	40	1	0.025	0.031
Year	1996	15–19	2151	271	0.126	0.123
n	6.135	20–24	3491	525	0.150	0.157
Mode	23.37	25–29	2822	452	0.160	0.150
s	0.825	30–34	1861	233	0.125	0.129
m	26.426	35–39	955	99	0.104	0.108
P	0.366	40–44	241	21	0.087	0.088
		45–49	37	3	0.081	0.072
Year	1997	15–19	2026	257	0.127	0.127
n	5.031	20–24	3596	708	0.197	0.197
Mode	23.497	25–29	2918	531	0.182	0.182
s	0.618	30–34	1749	254	0.145	0.141
m	19.76	35–39	915	87	0.095	0.102
P	0.884	40–44	195	15	0.077	0.072
		45–49	26	2	0.077	0.051
Year	1998	15–19	874	184	0.211	0.202
n	7.442	20–24	1852	483	0.261	0.277
Mode	22.765	25–29	1634	440	0.269	0.248
s	0.671	30–34	1484	283	0.191	0.194
m	20.029	35–39	921	123	0.134	0.144
P	0.089	40–44	271	28	0.103	0.105
		45–49	37	4	0.108	0.076
Year	1999	15–19	2067	341	0.165	0.163
n	7.880	20–24	2695	690	0.256	0.262
Mode	24.449	25–29	2186	577	0.264	0.257
s	0.654	30–34	1223	265	0.217	0.214
m	22.152	35–39	1183	192	0.162	0.166
P	0.810	40–44	322	39	0.121	0.125
		45–49	43	3	0.070	0.093

Urban population survey and migrant mineworkers

Carletonville, the biggest gold mining complex in the world, is situated about 100 km to the southwest of Johannesburg. The data presented below were collected as part of a broader survey to investigate not only HIV but also other sexually transmitted diseases as well as a range of social and behavioural determinants of the spread of infection. This project is discussed in detail in a separate paper in this issue.[5]

Sex workers

The Medical Research Council in Durban has been conducting research with sex workers at truck-stops along the national road linking Durban to Johannesburg since 1992. Between August 1996 and June 1998, 477 sex workers from five of the nine truck-stops in the KwaZulu-Natal Midlands were screened for HIV as a prerequisite for participation in a microbicide trial. There are an estimated 800 sex workers operating at the truck-stops in the Midlands. Participatory research methods were used to recruit women for the study. Clinical investigations were performed to identify sexually transmitted diseases and HIV status at baseline. Further details regarding the research methodology are described in Ramjee *et al.*[12]

Truck drivers

A seroprevalence survey was conducted among truck driver clients of sex workers, discussed above, at the truck-stops in the KwaZulu-Natal Midlands during 1998. Sex workers were asked to recruit participants from among their clients and a saliva test was used to determine HIV status.[11]

Workplace survey

An anonymous cross-sectional survey was carried out among the workforce of a major South African company to estimate the proportion of employees that are infected with HIV and to determine risk factors for HIV infection. Of all of the sites in the country with more than one hundred employees, 50% were chosen for inclusion in the study using probability-proportional-to-size sampling methods. A set number of employees was then recruited at each site. The results of this study were used to make projections on future HIV prevalence levels, to estimate the economic and other impacts of HIV on the company and to provide a baseline against which to evaluate the company's HIV/AIDS programme. This study provided information on overall HIV levels among men and women from all the provinces in South Africa.

Table 2. Age-prevalence of infection among women attending ante-natal clinics in Hlabisa,[6] KwaZulu-Natal. The first two columns give the parameters of the log-normal fits and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
Year	1997	10–14	7	0	0.000	0.010
n	8.140	15–19	1097	255	0.232	0.237
Mode	22.759	20–24	1491	530	0.355	0.345
s	0.612	25–29	1005	286	0.285	0.302
m	18.555	30–34	632	142	0.225	0.224
P	0.543	35–39	354	61	0.172	0.157
		40–44	113	10	0.088	0.107
		45–49	26	3	0.115	0.073
		50–54	4	1	0.250	0.050
		54–59	1	0	0.000	0.034
Year	1998	10–14	7	0	0.000	0.004
n	9.217	15–19	819	173	0.211	0.216
Mode	23.977	20–24	994	391	0.393	0.384
s	0.571	25–29	608	221	0.363	0.362
m	359	30–34	398	93	0.234	0.277
P	210	35–39	265	61	0.230	0.194
		40–44	57	7	0.123	0.131
		45–49	15	2	0.133	0.088
		54–59	3	1	0.333	0.039

Cancer patients at Chris Hani Baragwanath hospital

The HIV results were taken from two epidemiological studies, the first carried out between 1992 and 1994 and the second between 1995 and 1998. The study subjects included 283 men and 532 women admitted to Chris Hani Baragwanath, Hillbrow or Johannesburg hospitals for the first time with a cancer. Patients with cancer of the cervix, vulva, Kaposi's sarcoma and non-Hodgkin lymphoma, all found to be associated with HIV, were excluded.[13,14]

Results

Ante-natal clinic data

The national ante-natal clinic data are used in the previous section to illustrate the method of analysis. From those data the shape of the ante-natal age-prevalence curves for HIV infection appear to have remained much the same over a period of four years (excluding the results for 1996, which appear to be aberrant), except that in 1999 there seems to have been a shift to greater ages by about 1.5 years from an average of 22.9 years for the earlier data to 24.5 years for the 1999 data (Fig. 2, Table 1). This cannot be, since even if transmission stopped completely and no-one died of AIDS or AIDS-related diseases (or if the proportion who died was independent of age), the curve would shift up by only one year. It seems more likely, therefore, that there are differences in the sampling procedures or other biases in the data between the earlier years and 1999. While the observed changes may reflect saturation in the epidemic, one would then expect to see an increase in the scale parameter rather than an upward shift in the mode.

Ante-natal clinic data are also available for Hlabisa.[4,6] The shapes of the age-prevalence curves (Fig. 3) are similar to those for the national ante-natal clinic surveys with the peak prevalence, averaged over 1997 and 1998, occurring at 23.4 years (which compares with 22.9 years for the national surveys) and the scale parameter being only 8% smaller (Tables 1, 2).

Population surveys

Two sets of data from population surveys are available; one for a rural population in northern KwaZulu-Natal, carried out between 1990 and 1992;[9,10] and the second for the urban population of Carletonville, conducted in 1998 and discussed elsewhere in this issue.[5] The age-prevalences for these sets of data are shown in Fig. 4 using the results given in Table 3. As expected, the infection rates in the later urban survey are approximately 10 times those measured in the earlier rural survey (as a result of which the relative errors in the latter dataset are greater than in the former, even though the sample size was much greater). Nevertheless, the patterns of infection are similar between the two sets of data with infection rates increasing more slowly among young men (Fig. 4a, c) than among young women (Fig. 4b, d) and the peak prevalence occurs at greater ages among men than among women.

Migrant men

Two sets of data are available for migrant men: one for mineworkers and the other for truck drivers (Fig. 5; Table 4). For neither group does the age-prevalence deviate significantly from a constant, in contrast to all the other datasets, although the prevalence amongst the truck drivers is approximately twice that amongst the mineworkers.

Commercial sex workers

The age-prevalence of infection for commercial sex workers operating out of truck-stops in the KwaZulu-Natal Midlands is given in Fig. 6 and Table 5. For the women working at truck-stops, the data differ significantly from a straight line (χ^2

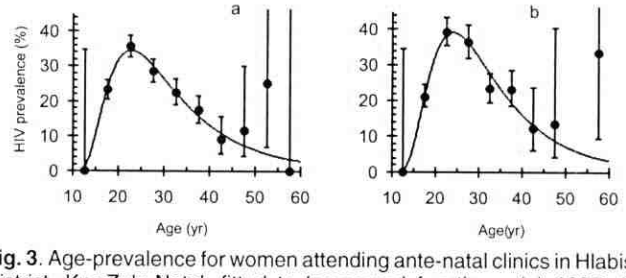

Fig. 3. Age-prevalence for women attending ante-natal clinics in Hlabisa District, KwaZulu-Natal, fitted to log-normal functions. (**a**) 1997; (**b**) 1998. The data, the parameters of the fits and the goodness of fit are given in Table 2. Error bars are 95% binomial confidence limits.

195.XLS

Table 3. Age-prevalence of infection among rural men and women[9,10] followed by the data for urban men and women.[5] The first two columns give the parameters of the log-normal fits and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
Year	1991	15–19	643	1	0.002	0.002
n	0.993	20–24	385	9	0.023	0.023
Mode	29.871	25–29	305	16	0.052	0.045
s	0.389	30–39	430	14	0.033	0.040
m	23.121	40–49	299	6	0.020	0.016
P	0.720	50+	354	2	0.006	0.005
Year	1991	15–19	1095	45	0.041	0.040
n	1.436	20–24	1088	50	0.046	0.049
Mode	21.926	25–29	739	34	0.046	0.043
s	0.745	30–39	1026	31	0.030	0.030
m	20.768	40–49	600	10	0.017	0.017
P	0.937	50+	624	6	0.010	0.010
Year	1998	13–14	31	0	0.000	0.000
n	10.939	15–19	99	2	0.020	0.017
Mode	31.646	20–24	89	15	0.169	0.186
s	0.416	25–29	59	25	0.424	0.390
m	25.727	30–34	49	21	0.429	0.443
P	0.962	35–39	60	22	0.367	0.377
		40–44	42	13	0.310	0.276
		45–49	35	5	0.143	0.186
		50–54	20	2	0.100	0.119
		54–59	15	2	0.133	0.074
Year	1998	13–14	30	0	0.000	0.022
n	14.898	15–19	101	21	0.208	0.215
Mode	26.173	20–24	119	64	0.538	0.518
s	0.546	25–29	117	68	0.581	0.573
m	21.799	30–34	105	49	0.467	0.483
P	0.554	35–39	93	31	0.333	0.361
		40–44	68	16	0.235	0.256
		45–49	30	6	0.200	0.177
		54–59	27	4	0.148	0.121

test of deviance, $P = 0.0013$) and the log-normal function gives a statistically good fit. For the women operating near a gold mine, the data do not differ significantly from a straight line but the sample size is small and the power of the test correspondingly weak. The rates among women operating near the mine are significantly higher than among the women at the truck-stops.

Workplace surveys

Figure 7 gives the age-prevalence of infection from a workplace survey carried out in a large company across the nation. There were few women in the survey and the data for men are separated into those who live in hostels or camps and those who reside in their own homes. For the men who live in hostels and camps, the age-prevalence peaks at about 35 years as it does for the population-based survey of men. For men living in their own homes, the prevalence is low and the sample size small, so that the error bars are relatively large and one can say only that the average prevalence among these men is about one third of that among the men who live in hostels or camps.

Cancer patients

The data for cancer patients include few young people and the estimates of infection rates in young adults is correspondingly poor but it does include a significant number of older people. In particular it provides a useful indication of the rates of infection among people over the age of 50 years. The data are given in Fig. 8 and Table 7, from which it can bee seen that the infection rates remain significant among both men and women above the age of 50 years. Among men aged 50 to 59 years, 7.8% (3.9–15.3%; 95% CL) are HIV positive; among women the corresponding figure is 11.5% (7.4–17.5%; 95% CL). Among men aged 60 to 69 years

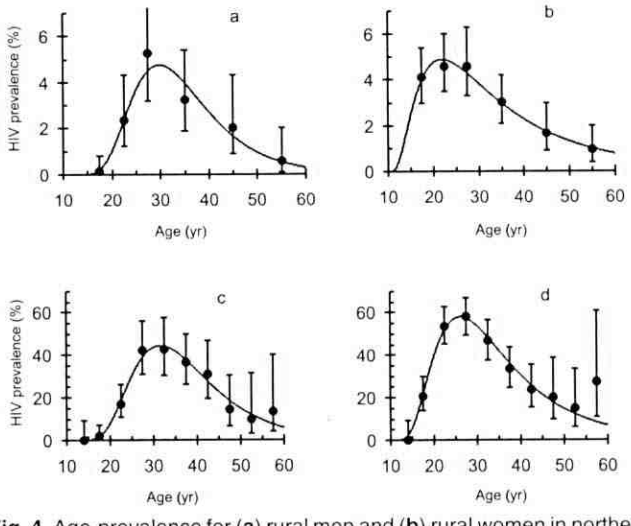

Fig. 4. Age-prevalence for (**a**) rural men and (**b**) rural women in northern KwaZulu-Natal in 1991 and (**c**) urban men and (**d**) urban women in Carletonville in 1998. The data, the parameters of the fits and the goodness of fit are given in Table 3. Error bars are 95% binomial confidence limits. Note the different vertical scales for figures (a) and (b), and (c) and (d).

196. RCS

Table 4. Age-prevalence of infection among mineworkers, above,[5] and truck drivers, below.[11] The first two columns give the parameters of the linear fits and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
Year	1998	20–24	82	24	0.293	0.286
Constant	0.286	25–29	167	47	0.281	0.286
P	0.506	30–34	191	54	0.283	0.286
		35–39	196	62	0.316	0.286
		40–44	129	36	0.279	0.286
		45–49	83	24	0.289	0.286
		50–54	38	9	0.237	0.286
		54–59	11	1	0.091	0.286
Year	1998	20–24	23	11	0.478	0.562
		25–29	61	31	0.508	0.562
Constant	0.562	30–34	56	32	0.571	0.562
P	0.502	35–39	58	29	0.500	0.562
		40–44	43	26	0.605	0.562
		45–49	44	29	0.659	0.562
		50–54	22	13	0.591	0.562

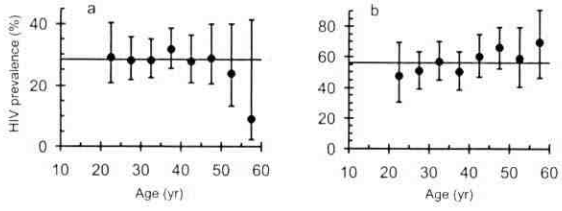

Fig. 5. Age-prevalence for (**a**) mineworkers and (**b**) truck drivers, both in 1998. The data, the parameters of the fits and the goodness of fit are given in Table 3. Error bars are 95% binomial confidence limits. Note the different vertical scales for figures (a) and (b).

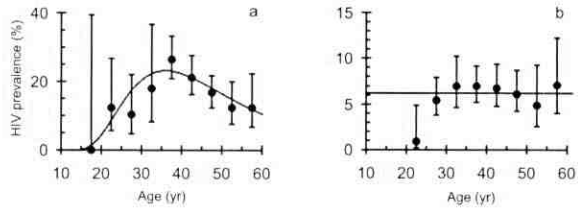

Fig. 7. Age-prevalence for men working in a large parastatal industry. (**a**) Men who live in a hostel or a camp, (**b**) men who live in their own home both in 1999. The data, the parameters of the fits and the goodness of fit are given in Table 6. Error bars are 95% binomial confidence limits.

the rate of infection is still high at 3.9% (1.4–10.9%; 95% CL), while among women it has dropped to 1.1% (0.3–3.9%; 95% CL).

Discussion

The data presented above show four main patterns of infection: among women attending ante-natal clinics, among women in population surveys, among men in population surveys, and finally among migrant men. What is striking is that the pattern of infection does not appear to be changing substantially with time (although as the epidemic saturates it must) and does not appear to differ between urban and rural populations (see Fig. 4). These observations are illustrated in Fig. 9, where the 95% confidence ellipses are plotted for the most precise datasets described above. The data for women attending ante-natal clinics are tightly clustered, showing that the shapes of the curves are similar not only among years but also between the national surveys and the rural survey. The data for urban women, on the other hand, show a slightly narrower distribution (smaller value of σ) but the peak prevalence occurs at a significantly greater age (higher value of the mode). The data for urban men show a significantly narrower distribution than do the results for urban women and the maximum prevalence occurs 5.4 years later.

To illustrate the above observations in more detail, Fig. 10 shows a series of comparisons in which the curves are all scaled so that the area under each curve is 1. This is done to facilitate the comparisons of the shapes of the curves, since the average prevalence varies widely among the datasets. Figure 10a shows first a comparison of the age-specific incidence of pregnancy and the average fit to the age-specific prevalence of HIV among women attending ante-natal clinics in the years 1995, 1997 and 1998. Because the age-specific incidence of pregnancy gives a measure of sexual activity, it also provides a measure of risk of infection, although this will be confounded by the use of non-barrier methods of contraception. Nevertheless, the curves are surprisingly similar. In Fig. 10b the data for women attending ante-natal clinics are compared to those for urban women. The ante-natal clinic data overestimate infection rates in younger women and underestimate these rates in older women as compared to the urban population survey. The best comparison

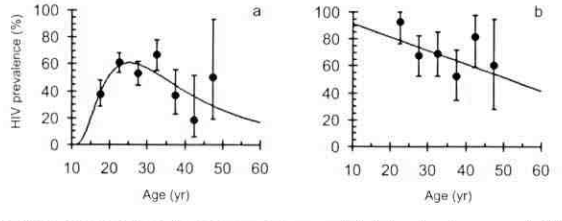

Fig. 6. Age-prevalence for sex workers at (**a**) truck-stops and (**b**) at a mine, both in 1998. The data, the parameters of the fits and the goodness of fit are given in Table 5.

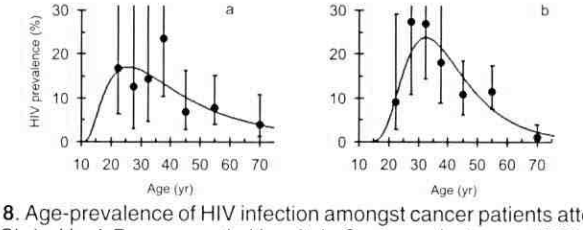

Fig. 8. Age-prevalence of HIV infection amongst cancer patients attending Chris Hani Baragwanath Hospital, Gauteng, between 1992 and 1994. (**a**) Men and (**b**) women. The data, the parameters of the fits and the goodness of fit are given in Table 7. Error bars are 95% binomial confidence limits.

198.XLS

Table 6. Age-prevalence of infection among men working for a large parastatal. (**a**) Those who own their own homes and (**b**) those who live in hostels or camps. The first two columns give the parameters of the log-normal fits (a) and the linear fits (b) and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
(**a**)						
Year	1998	15–19	6	0	0.000	0.393
n	9.009	20–24	40	5	0.068	0.143
Mode	35.976	25–29	49	5	0.056	0.120
s	0.521	30–34	28	5	0.097	0.189
m	34.076	35–39	196	52	0.056	0.067
P	0.477	40–44	193	41	0.051	0.064
		45–49	246	41	0.042	0.052
		50–54	113	14	0.048	0.075
		55–59	72	9	0.057	0.099
		60–65	33	5	0.152	0.093
(**b**)						
Year	1998	20–24	112	1	0.009	0.062
Constant	0.062	25–29	459	25	0.054	0.062
P	0.110	30–34	317	22	0.069	0.062
		35–39	643	45	0.070	0.062
		40–44	448	30	0.067	0.062
		45–49	458	28	0.061	0.062
		50–54	165	8	0.048	0.062
		55–59	156	11	0.071	0.062
		60–65	37	3	0.081	0.062

Table 7. Age-prevalence of infection amongst cancer patients attending Chris Hani Baragwanath Hospital, Gauteng, between 1995 and 1998. (**a**) Men and (**b**) women. The first two columns give the parameters of the log-normal fits (a) and the linear fits (b) and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
(**a**)						
Year	1995–98	20–24	18	3	0.167	0.163
n	7.130	25–29	8	1	0.125	0.170
Mode	25.900	30–34	14	2	0.143	0.155
s	0.775	35–39	17	4	0.235	0.133
m	28.998	40–49	59	4	0.068	0.102
P	0.797	50–59	90	7	0.078	0.069
		60–79	77	3	0.039	0.039
(**b**)						
Year	1995–98	20–24	22	2	0.091	0.104
n	6.783	25–29	11	3	0.273	0.205
Mode	32.492	30–34	26	7	0.269	0.238
s	0.456	35–39	33	6	0.182	0.216
m	27.680	40–49	101	11	0.109	0.149
P	0.240	50–59	157	18	0.115	0.075
		60–79	182	2	0.011	0.023

1. Department of Health (2000). *National HIV Seroprevalence Survey of Women Attending Public Antenatal Clinics in South Africa*. Department of Health, Pretoria.
2. Williams B.G. and Campbell C.M. (1998). Understanding the epidemic of HIV in South Africa: analysis of the antenatal clinic survey data. *S. Afr. Med. J.* **88**, 247–251.
3. Janssen R.S., Satten G.A., Stramer S.L., Rawal B.D., O'Brien T.R., Weiblen B.J., Hecht F.M., Jack N., Cleghorn F.R., Kahn J.O., Chesney M.A. and Busch M.P. (1998). New testing strategy to detect early HIV-1 infections for use in incidence estimates and for clinical and prevention purposes. *J. Am. Med. Assoc.* **280**, 42–48.
4. Williams B., Gouws E., Wilkinson D. and Abdool Karim S.S. (submitted). Estimating HIV incidence rates from age prevalence data in epidemic situations.
5. Williams B.G., MacPhail C., Campbell C., Taljaard D., Gouws E., Moema S., Mzaidume Y. and Rasego B. (2000). The Carletonville-Mothusimpilo project: limiting transmission of HIV through community participation. *S. Afr. J. Sci.* **96**, 351–359.
6. Wilkinson D., Abdool Karim S.S., Williams B.G. and Gouws E. (in press). High HIV incidence and prevalence among young women in rural South Africa: developing a cohort for intervention trials. *Journal of Acquired Immune Deficiency Syndromes*.
7. Williams B.G. and Dye C. (1994). Maximum likelihood for parasitologists. *Parasitology Today* **10**, 489–493.
8. Department of Health (1999). *1998 National HIV Sero-prevalence Survey of Women Attending Antenatal Clinics in South Africa*. Department of Health, Pretoria.
9. Abdool Karim Q., Abdool Karim S.S., Singh B., Short R. and Ngxongo S. (1992). Seroprevalence of HIV infection in rural South Africa. *AIDS* **6**, 1535–1539.
10. Abdool Karim Q. (2000). *Women and AIDS: Epidemiology and Gender Barriers to Prevention in KwaZulu-Natal, South Africa*. Ph.D. thesis, University of Natal, Durban.
11. Ramjee G., Gouws E., Stein Z. and Nyembe M. (2000). HIV prevalence among truck drivers in KwaZulu-Natal, South Africa — implications for the explosive nature of the South African HIV epidemic. *International AIDS Conference, Durban 2000*. Abstract no. D1194
12. Ramjee G., Abdool Karim S.S. and Sturm A.W. (1998). Sexually transmitted infections among sex workers in KwaZulu-Natal, South Africa. *Sexually Transmitted Diseases* 25, 346–349
13. Sitas F, Bezwoda W.R., Levin V. *et al.* (1997). Association between HIV-1 infection and cancer in the black population of Johannesburg and Soweto. *Br. J. Cancer* **5**, 1704–1707.
14. Sitas F., Pacella-Norman R., Carrara H. *et al.* (in press). The spectrum of HIV-related cancers in South Africa. *International Journal of Cancer*.

197. XLS

Table 5. Age-prevalence of infection among commercial sex workers[5] operating at truck-stops[11,12] (above), at a gold mine[5] (below). The first two columns give the parameters of the log-normal fits (a) and linear fits (b) and their values, the remaining columns the age ranges, the numbers sampled and HIV positive, and the observed and fitted prevalences.

Parameter	Value	Age class	Total	Positive	Prevalence	Fit
Year	1998	15–19	31	12	0.387	0.359
n	22.541	20–24	56	33	0.589	0.618
Mode	25.212	25–29	41	28	0.683	0.684
s	0.740	30–34	41	29	0.707	0.646
m	26.301	35–39	11	5	0.455	0.570
P	0.029	40–44	2	0	0.000	0.485
		45–49	4	2	0.500	0.406
Constant	1.014	25–29	14	13	0.929	0.789
Slope	–0.010	30–34	34	23	0.676	0.739
P	0.074	35–39	26	18	0.692	0.689
		40–44	25	13	0.520	0.639
		45–49	11	9	0.818	0.589
		50–54	5	3	0.600	0.539

between men and women is based on the urban population survey carried out in Carletonville in 1998 and these data are shown in Fig. 10c, from which it is seen that the curve for men is shifted to older age-groups, as compared to women in the same community, by about five years. Finally, Fig. 10d shows a comparison of the fits to the data for urban men and migrant workers, with the latter being essentially constant with age.

A number of important conclusions follow from this analysis. First, when the epidemic growth rate is very high, the age-specific prevalence of infection gives an approximate measure of the age-specific incidence of infection and this is supported by the fact that the shapes of these curves for particular populations do not vary greatly over time. This notion is supported by the similarity between the age-specific incidence of pregnancy and the age-specific prevalence of infection among women attending ante-natal clinics. Reliable estimates of the age-specific incidence of infection are probably the most important variables needed to make demographic models to forecast the course of the epidemic and against which to assess the impact of interventions.

Comparing the age-specific prevalence of infection among women attending ante-natal clinics with that among urban women shows that the ante-natal clinic data tend to overestimate the prevalence of infection among younger women and underestimate it among older women.

The age-specific prevalence of infection among men is shifted up by about five years as compared to that for women, probably because males are usually older than their female sexual partners.

The two datasets for migrant workers show a completely different pattern of infection with no significant variation with age. There are no very young migrant workers in the sample but these data suggest that among older migrant workers the risk of infection does not fall off among older men in the way it does among all of the other groups for which data have been presented. It seems reasonable to suppose that migrant men, through the nature of their work, remain at high risk of infection as they get older, because they do not have the social support provided in even the poorest of stable communities.

Finally, it is of interest to note that there is no significant difference between the datasets for people living in urban and rural areas as indicated by the results shown in Fig. 4, so that the age-risk of infection does not appear to depend on whether or not people live in urban or rural areas.

These data show alarming rates of increase in prevalence with age (and so also indicate very high incidences) in young women aged 15–20 years, with prevalences reaching up to 60% in some groups of 25-year-old women who are not commercial sex workers. Further studies are needed, as a matter of extreme urgency, both to understand the reasons for this and to find ways to protect young women from becoming infected. These data also highlight the increased risk of infection among migrant men, especially among older men, and this must be taken into account when designing interventions to manage the epidemic.

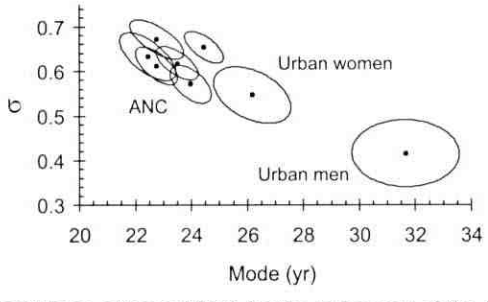

Fig. 9. Confidence ellipses (95%) for the parameter of the fits to the national ante-natal clinic (ANC) data for the years 1995, 1997, 1998 and 1999 and the rural ante-natal clinic data for 1987 and 1988; the data for urban women in 1998; and the data for urban men in 1998.

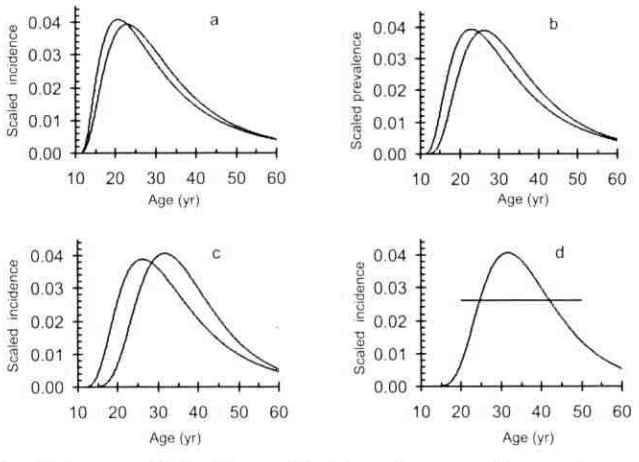

Fig. 10. Age-specific incidence of fertility and age-specific prevalence of HIV infection. (**a**) Left: fertility; right: women attending ante-natal clinics. (**b**) Left: women attending ante-natal clinics; right: urban women. (**c**) Left: urban women; right: urban men. (**d**) Curve: urban men; straight line: migrant men. All the curves are scaled so that the area under the curve is one. The line for the migrant men is scaled by the same factor as the curve for urban men.