PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Transmission Dynamics of Human Immunodeficiency Virus (HIV) [and Discussion]

R. M. May, R. M. Anderson and M. E. Irwin

Phil. Trans. R. Soc. Lond. B 1988 **321**, 565-607
doi: 10.1098/rstb.1988.0108

Phil. Trans. R. Soc. Lond. B **321**, 565–607 (1988)
Printed in Great Britain

The transmission dynamics of human immunodeficiency virus (HIV)

By R. M. May,[1] F.R.S., and R. M. Anderson,[2] F.R.S.

[1] *Department of Biology, Princeton University, Princeton, N.J. 08544, U.S.A.*
[2] *Department of Pure and Applied Biology, Imperial College, London University, London SW7 2BB, U.K.*

The paper first reviews data on HIV infections and AIDS disease among homosexual men, heterosexuals, intravenous (IV) drug abusers and children born to infected mothers, in both developed and developing countries. We survey such information as is currently available about the distribution of incubation times that elapse between HIV infection and the appearance of AIDS, about the fraction of those infected with HIV who eventually go on to develop AIDS, about time-dependent patterns of infectiousness and about distributions of rates of acquiring new sexual or needle-sharing partners.

With this information, models for the transmission dynamics of HIV are developed, beginning with deliberately oversimplified models and progressing – on the basis of the understanding thus gained – to more complex ones. Where possible, estimates of the model's parameters are derived from the epidemiological data, and predictions are compared with observed trends. We also combine these epidemiological models with demographic considerations to assess the effects that heterosexually-transmitted HIV/AIDS may eventually have on rates of population growth, on age profiles and on associated economic and social indicators, in African and other countries. The degree to which sexual or other habits must change to bring the 'basic reproductive rate', R_0, of HIV infections below unity is discussed. We conclude by outlining some research needs, both in the refinement and development of models and in the collection of epidemiological data.

1. Introduction

As is abundantly recorded in prose and picture throughout history, plague and pestilence have always excited more dread than the other three horsemen of the Apocalypse. Infectious diseases are frightening not simply because they can kill or cripple, but because the process of infection seems so insidious. Other chance events, such as car accidents (which in developed countries kill many more people each year than does AIDS, although this could change in the U.S.A. in a few years), give the illusion of being under human control. But the inapparent nature of most infectious agents seemingly puts them beyond control, and gives rise to anxieties that are affected little by knowing, for example, that AIDS is caused by a virus whose molecular sequence is known, rather than by some medieval miasma.

The Human Immunodeficiency Virus (HIV), the causative agent of Acquired Immuno-Deficiency Syndrome (AIDS), has several properties which make it especially liable to arouse such anxieties. For one thing, a high proportion of HIV infections (30% or more) lead to the disease AIDS and thus to death, with no cure and no vaccine in sight. For another, HIV infection as such is effectively asymptomatic in most cases, and the incubation interval to develop AIDS-Related Complex (ARC) or AIDS is long and variable. Moreover, not enough is yet known about basic epidemiological parameters to make reliable long-term predictions about likely patterns

of infection among particular groups (heterosexuals, homosexual males, IV-drug abusers, and so on) in particular countries or regions.

The epidemic spread or endemic maintenance of an infection depends on its basic reproductive rate, R_0, within the host population in question. For a 'microparasitic' infection (*sensu* Anderson & May 1979) such as HIV, R_0 is defined as the average number of secondary infections produced by one primary infection within a wholly susceptible population (Ross 1911; Macdonald 1956; Smith 1970; Anderson & May 1979, 1985). If R_0 exceeds unity, an epidemic will on average spread, expanding along chains of transmission in which each infection produces more than one 'offspring'; if R_0 is less than unity, the infection dies out following each introduction. As discussed more rigorously below, for HIV/AIDS and other sexually transmitted diseases R_0 is found, from its definition, by multiplying together the probability of infecting any one partner, the average number of new partners per unit time and the average duration of infectiousness (which may, or may not, be the average time from acquiring HIV to dying from AIDS). Each of these three basic components – incubation period and pattern of infectiousness, rates of acquiring new partners of a specified kind, and transmission probabilities – is currently ill-understood for HIV/AIDS.

Our paper deals with these and other aspects of the transmission dynamics of HIV/AIDS. We first survey and summarize what is currently known about the epidemiological processes outlined above. We then present a range of simple mathematical models designed to capture the essentials of these processes, and we indicate what can be learned from the models. A further section combines epidemiological with demographic considerations, to explore the possible effects of HIV/AIDS on age-profiles and other social and economic characteristics in countries such as Africa where the disease seems to be widely disseminated. We conclude by discussing the use of such models to indicate the kinds of data needed for long-term assessment, and as points of departure for the numerical exploration of more detailed and more realistic models (as appropriate data become available). We also indicate how such models can be used to make inferences about individual-level parameters (such as transmission probabilities) from population-level data (such as the way seroprevalence changes over time).

2. Basic epidemiological factors

2.1. *Groups at risk for HIV/AIDS*

In sub-Saharan Africa, HIV/AIDS appears to be mainly transmitted by sexual contacts among heterosexuals (Quinn *et al.* 1986). In developed countries, homosexual and bisexual males account for most (around 65–85% or more) cases of AIDS. IV-drug abusers also constitute a significant fraction of AIDS cases in developed countries, with significant variations within and between countries; IV-drug abusers make up about half the recently reported AIDS cases in northeastern regions of the U.S.A.

HIV infections and AIDS cases resulting from blood transfusions in general, and among haemophiliacs in particular, amounted to a small fraction of the total in earlier years. In developed countries there are now very few, if any, new infections in these categories, following the introduction of screening and testing of blood supplies. This transmission route is still open in some developing countries, however, and its role in the overall transmission pattern in such countries remains somewhat uncertain.

HIV/AIDS can be transmitted vertically to the offspring of infected mothers, and these

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

'paediatric AIDS' cases are increasing whenever AIDS is reported. Although very uncertain, current estimates are that perhaps 30–50% or more of the children born to infected mothers will die of AIDS in their first few years. It is conceivable that HIV infection may also be transmitted vertically from males, in their sperm. We are not aware of any discussion of this possibility, but if it exists it could complicate the analysis of data, as well as adding to the significance of vertical transmission.

The age-specific incidence of AIDS cases in both developed and developing countries seems to rule out any significant transmission of HIV infection by insect vectors or by contaminated needles in public health programmes. Were it otherwise, one would expect to see more cases in the 5–15-year-old age-range (although it is conceivable that multiple use of vaccination needles in developing countries is concentrated within age-cohorts, thus not propagating HIV among children so long as the prevalence remains very low in this age group; there is room for further empirical and theoretical studies of this question).

These facts lead us initially to study models for the transmission of HIV/AIDS among homosexual males in large cities in the U.K., U.S.A. and other developed countries, and among heterosexuals in Africa. Similar models pertain to transmission among IV-drug abusers. Public health planning will eventually require detailed studies involving more complex models, embracing all the transmission routes discussed above and assigning magnitudes to the parameters that characterize the linkages among different categories (Hethcote 1987); little is known about these parameters.

The different patterns in different countries are reflected in ratios of AIDS cases among men to cases among women. The current overall incidence of AIDS in the U.S.A is 13.0 times greater for males than for females (HHS 1987), which accords with the ratios of 14–20:1 found in European countries (Anderson *et al.* 1986; May & Anderson 1987). As summarized in table 1, however, studies among particular groups in the U.S.A. give lower ratios, which seems reasonable given the nature of the groups. In Africa, by contrast, the sex ratio of AIDS cases runs around 1:1, with perhaps a slight excess of female cases (Quinn *et al.* 1986).

TABLE 1. RATIO OF AIDS CASES OR HIV PREVALENCE AMONG MALES TO THAT AMONG FEMALES, FOR SOME PARTICULAR STUDIES IN THE U.S.A. (FROM HHS 1987)

group	statistic	ratio of males to females
all AIDS cases	cumulative number of reported AIDS cases, through 1987	13.0:1
heterosexual adult and adolescent AIDS cases	cumulative number assigned to this category, CDC data through 1987	2.9:1
military recruit applicants	seroprevalence of HIV; data from Department of Defense, 1985–87	5.5:1
American Red Cross blood donors	nationwide seroprevalence of HIV, 1986–87:	
	first time donors	4.6:1
	repeat donors	4.6:1
sentinel hospital patients in Midwest	seroprevalence of HIV, unpublished CDC data	2.3:1
IV-drug abusers	New York City, 1985	0.9:1
IV-drug abusers	four cities in Connecticut, 1986–87	1.2:1

2.2. *Prevalence of HIV infection and AIDS cases*

As of November 1987, 126 countries had reported at least one case of AIDS. In the U.S.A. around 55000 cases of AIDS (resulting so far in 31000 deaths) had been reported by March 1988. The corresponding numbers for the U.K. are around 1200 cases and 700 deaths, as of January 1988. Figure 1 shows temporal trends in the numbers of AIDS cases for several countries.

Although complicated by lags in reporting and changes in the definition of what constitutes AIDS (Brookmeyer 1988), the data for AIDS cases are much more abundant than those for HIV infections. HIV infections are usually asymptomatic, and many of those infected are unaware

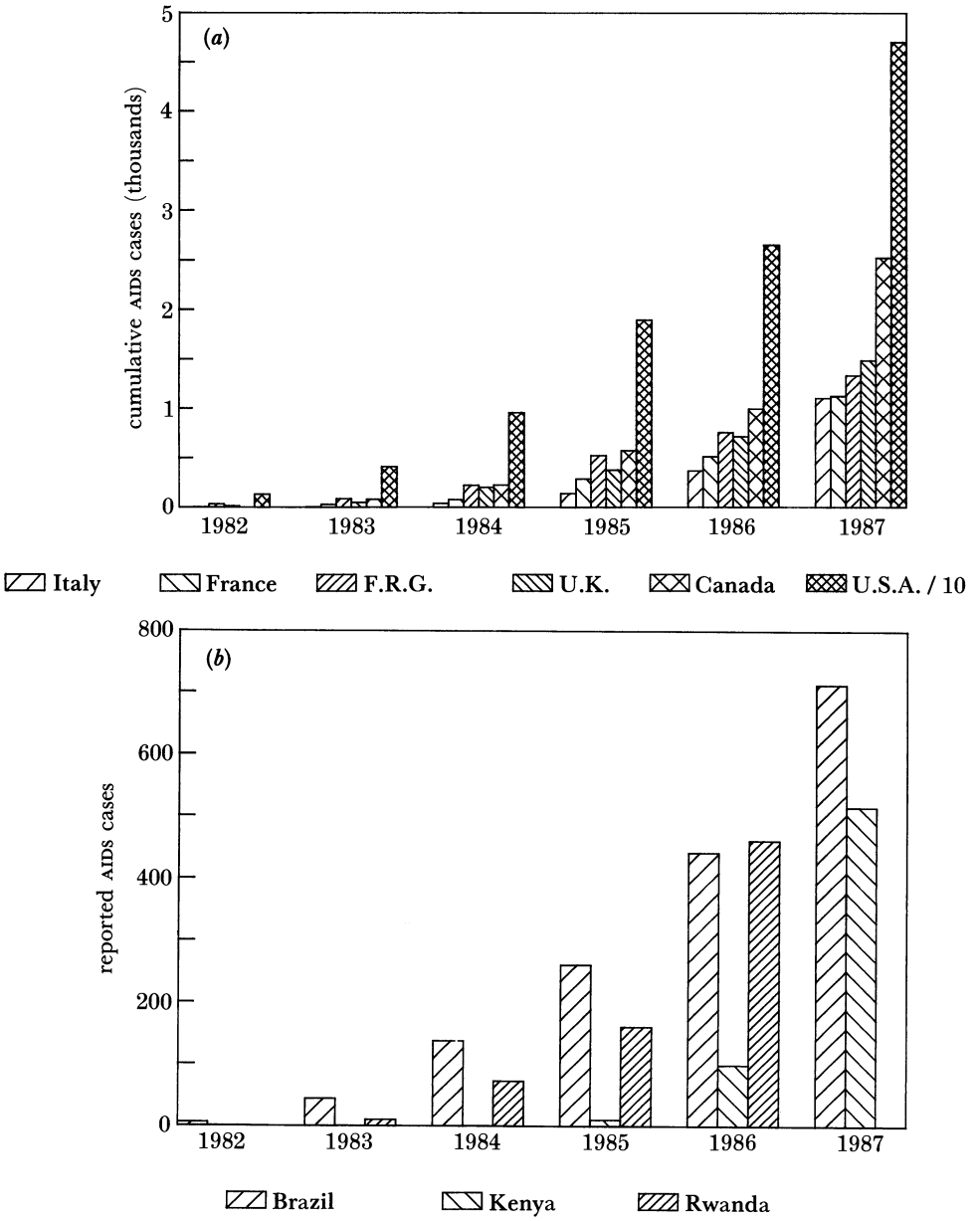

FIGURE 1. Cumulative number of cases of AIDS in (*a*) some European and North American countries and (*b*) in Brazil, Kenya and Rwanda (as reported to the World Health Organization).

of their state. Moreover, there is no registry for compiling data of HIV infections, and even if there were many of those infected would actively avoid such a registry. In many countries, plans for surveys of HIV seropositivity among unbiased samples of the population are being held up by worries about the feasibility of obtaining essentially 100% compliance (or at least unbiased non-compliance), by ethical dilemmas about how to counsel those found to be seropositive, and by technical problems concerning false positives and negatives. Thus the kinds of studies available all consequently involve groups that are biased in one way or another: blood transfusion recipients; attendees at clinics for sexually transmitted diseases (STDs); pregnant women; military personnel or applicants for the armed services; female prostitutes; and others. Figure 2 shows data for the change in seroprevalence over time for some of these groups.

In the U.S.A. as of November 1987 some 50 studies of homosexual and bisexual men, in different regions, show seroprevalence levels ranging from under 10% to as high as 70%, with most findings between 20% and 50%. Among IV-drug abusers, 88 surveys and studies find seroprevalence ranging from highs of 50–65% in the vicinity of New York City and in Puerto Rico, to rates that – although variable – are mostly below 5% in areas other than the east coast. For individuals requiring treatment with clotting factor concentrates, seroprevalence is around 70% for haemophilia A and around 35% for haemophilia B; the rates are uniform throughout the U.S.A., reflecting the lack of any significant geographical factor in the distribution of haemophilia. Other special groups include prisoners and female prostitutes. Some 33 studies of HIV prevalence among prisoners showed levels ranging from 0 to 17%, which – as expected – was higher than among the general population but lower than among high-risk groups in the same general region. Some 19 studies of female prostitutes showed levels ranging from 0 to 45%, with the highest rates seen in large inner-city areas where IV-drug abuse is common; HIV prevalence was 3 to 4 times higher among prostitutes who acknowledged IV-drug abuse than among those who did not (Centres for Disease Control 1987; Johnson 1988).

Among the American population as a whole, such data as are available suggest HIV seroprevalence levels of the general order of 10^{-3}. For instance, Red Cross blood donors who have not previously been tested currently average 0.04% seropositive; applicants for military service (whom the U.S. military believe to under-represent persons in high-risk groups, although we do not share their certainty) run around 0.15%; Job Corps entrants ('disadvantaged' 16–21-year-olds) average around 0.33%; patients without AIDS-like conditions at 4 hospitals have seroprevalence levels around 0.32%; childbearing women in Massachusetts (tested anonymously through filter-paper blood specimens from their newborn infants) ran around 0.21%; 27 surveys in women's health clinics (excluding drug users) produced levels ranging from 0 to 2.6% (HHS 1987). The overall figure accords with the independent estimate of around 1.5 million HIV infections within the U.S.A. population of around 240 million. This rough figure is arrived at by adding up the estimated numbers in various categories of risk, along with rough estimates of seroprevalence levels in the various categories; the estimated number of homosexual males, for example, is based on the old Kinsey study (Kinsey *et al.* 1948).

Attempts have been made to refine the above U.S.A. data by removing those contributions to overall seroprevalence made by individuals in defined risk-groups (homosexual and bisexual males, IV-drug abusers, and so on). This leads to the conclusion that 'HIV prevalence levels in persons without acknowledged or recognized risks would be below 0.021% in military

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

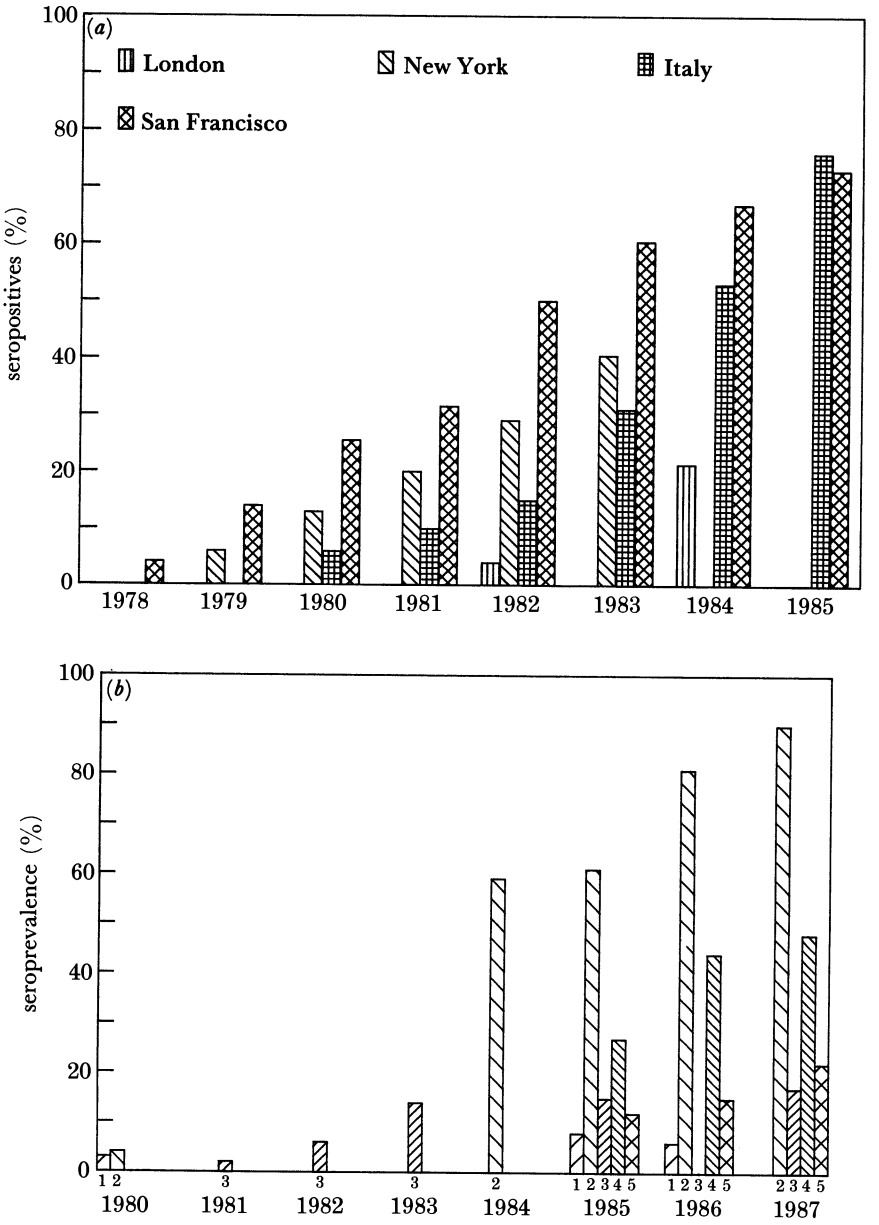

FIGURE 2. (*a*) The rise in seropositivity to HIV antigens in cohorts of patients over the period 1978–85. The studies in San Francisco, London and New York are of homosexual/bisexual males, and the study in Italy is of IV-drug abusers (for sources of these data, see May & Anderson (1987)). (*b*) Longitudinal changes in HIV seropositivity among people in various risk-groups in different countries in Africa: (1) pregnant women in Zaire; (2) female prostitutes in Nairobi, Kenya; (3) men with chancroid, Kenya; (4) men with STD infections, Rwanda; (5) pregnant women in Kinshasa, Zaire.

applicants and 0.006% in blood donors' (HHS 1987, p. 3). Limited studies of heterosexuals being treated for other STDs, but excluding those with identified risk factors by rigorous protocols of interviews, find seroprevalence levels ranging from 0 to 1.2% (in comparison with over 50% for homosexual males attending the same clinics).

In Africa, where transmission is thought to be mainly by heterosexual contacts, a different picture is emerging. In Nairobi, tests for HIV antibodies in blood sera stored from earlier studies

(often for different purposes) show seropositivity levels among female prostitutes rising from 4% in 1980–81 to around 51% in 1983–84 and 59% in 1985–86, and from 1% to 14% and 18% at the corresponding times for males attending a clinic for sexually transmitted diseases (Quinn *et al.* 1986). Recent seroprevalence levels among high-risk groups, such as female prostitutes, have been reported to range from 27% to 88%, depending on socioeconomic status and geographic location. In a general survey of pregnant women in Nairobi, seropositivity levels rose from 0% in 1980–81 to 2% in 1985–86; a similar survey of pregnant women in Kinshasa shows seropositivity levels rising from 0.2% in 1970 to 3% in 1980–81 and 8% in 1985–86 (Quinn *et al.* 1986). A study of some 600 seronegative men and women working in a general hospital in Kinshasa from 1984 to 1985 found the annual incidence (or rate of seroconversion) of HIV antibodies to be around 0.8%, a figure that Quinn *et al.* (1986) think may be representative of the annual incidence of new HIV infections in central and East Africa. Current studies of HIV seropositivity among blood donors in central African countries suggest levels as high as 9% in Zaire, 11% in Uganda, 15% in Rwanda, and 18% in Zambia (T. C. Quinn, personal communication).

In short, much more data are available about AIDS cases in developed regions than about HIV infections, and much more data for HIV infections among high-risk groups in developed countries that for non-drug-using heterosexuals in developed countries or for developing countries more generally.

Against this background, we now turn to examine the specific factors that influence the transmission dynamics of HIV.

2.3. *Incubation intervals and infectiousness*

One of the many things that makes HIV/AIDS different from most other infectious diseases is the long and variable incubation interval between infection with HIV and the collapse of the immune system that results in AIDS. Empirical information about the incubation interval, much less any fundamental understanding, is hard to obtain because it is not usually known when a given individual first acquired HIV infection. Most of the available data about incubation intervals come from transfusion-associated AIDS cases, where the date of the infection can usually be presumed to be the date of transfusion. Analysis of such data suggests an average incubation interval of around 8–9 years (Medley *et al.* 1987, 1988) (earlier analyses gave answers ranging from around 4 to around 15 years). These authors also found average incubation intervals to be significantly shorter for infants (around 2 years for those aged 1–4) and somewhat shorter for older people (around 5–6 years among those over 60), although this latter effect may be associated with the confounding influence of other sources of mortality. Medley *et al.* emphasize that the average incubation interval they have inferred is much the same as the longest incubation interval in the data, and thus should not be seen as graven in stone. They also emphasize that transfusion-associated infections may not be characteristic of HIV/AIDS acquired by other routes; there are, indeed, some tentative indications that average incubation intervals are somewhat shorter for homosexual males and IV-drug abusers.

The cohort studies of Medley *et al.* (1987, 1988) are consistent with analyses of stored serum samples, taken from homosexual men in San Francisco as part of a study of hepatitis B virus (HBV) that reaches back to 1978. In this serendipitous longitudinal study, it is possible to see when individuals seroconverted. The current figures are that, of those who have seroconverted, none had AIDS after 3 years, 20% had AIDS after 6 years, and currently 36% have AIDS after

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

7–8 years. Taken together, the current data suggest that around 30–40% of HIV infectees go on to develop AIDS. But there is evidence of immunological deterioration in 80% of those who have been infected for 8–9 years, and all those infected with HIV may well eventually go on to develop AIDS (unless they die from other causes) after characteristic incubation times that may be significantly greater than current estimates of the average period.

In general, the incidence of AIDS cases, the incidence of HIV infections and the distribution of incubation intervals are connected by

$$\mathrm{d}C(t)/\mathrm{d}t = \int_0^t I(s)\, d(t-s)\, \mathrm{d}s. \tag{2.1}$$

Here $C(t)$ is the cumulative number of AIDS cases up to time t since the infection first appeared (corrected for reporting lags and so on), so that $\mathrm{d}C/\mathrm{d}t$ is the rate at which new cases appear; $I(t)$ is the corresponding rate at which new HIV infections appear; and $d(t)$ is the distribution of incubation times (that is, $d(t)$ is the probability that someone who acquired HIV infection at $t = 0$ will develop AIDS at time t). Equation (2.1) is a standard Volterra integral equation of the first kind, and thus if complete information about any two of the quantities $C(t), I(t), d(t)$ is available then the third can be deduced. Specifically, the convolution (or 'faltung') theorem can be used to express I in terms of C and d as follows (Morse & Feshbach 1953, ch. 8):

$$I(t) = \frac{1}{2\pi \mathrm{i}} \oint_{\mathrm{C}} \frac{p\tilde{C}(p)\, \mathrm{e}^{pt}\, \mathrm{d}p}{\tilde{d}(p)}. \tag{2.2}$$

Here $\tilde{C}(p)$ and $\tilde{d}(p)$ are the Laplace transforms of $C(t)$ and $d(t)$, respectively, and the contour integral is over the standard Bromwich contour. The corresponding expression for $d(t)$ when $C(t)$ and $I(t)$ are known is obvious.

Unfortunately information about both the incidence of HIV infection, $I(t)$, and the distribution in incubation intervals, $d(t)$, is usually insufficient. In essentials, what Medley *et al.* (1987, 1988) do is use the available data about $C(t)$ and $I(t)$, along with assumptions about the functional shape of $d(t)$, to estimate the characteristic parameters of the distribution in incubation times. Conversely, if an explicit assumption is made about the incubation distribution, some functional shape can be assigned to $I(t)$ to estimate the characteristic parameters; in this way, the HHS (1987) report arrives at the independent estimate that the total number of HIV infections in the U.S.A. at the end of 1987 is in the approximate range 1–3 million (which is not inconsistent with the estimated 1.5 million mentioned earlier).

One simple (and epidemiologically conventional) assumption is that infected individuals go on to develop AIDS at some constant rate, v. The distribution of incubation intervals is then

$$d(t) = v \exp(-vt), \tag{2.3}$$

and the average incubation interval is $1/v$ (Anderson *et al.* 1986). More generally, suppose the probability $v(t)$ of developing AIDS increases as the time t since acquiring HIV infection lengthens, according to $v(t) = \alpha t^{\nu}$. The result is a Weibull distribution in incubation intervals,

$$d(t) = \alpha t^{\nu} \exp[-\alpha t^{\nu+1}/(\nu+1)]. \tag{2.4}$$

The available data are fitted reasonably well by such a Weibull distribution with $\nu = 1$, and this distribution is used in many of our numerical studies. The Weibull distribution gives an average incubation time of

$$\left(\frac{\nu+1}{\alpha}\right)^{1/(\nu+1)} \Gamma\left(\frac{\nu+2}{\nu+1}\right),$$

and the coefficient of variation of the distribution is

$$CV = \left[\Gamma\left(\frac{\nu+3}{\nu+1}\right) \Big/ \Gamma^2\left(\frac{\nu+2}{\nu+1}\right)\right] - 1$$

(for $\nu = 0$, the mean is $1/\alpha$ and $CV = 1$; for $\nu = 1$, the mean is $(\pi/2\alpha)^{\frac{1}{2}}$ and $CV = 0.27\ldots$).

The simplest assumption is that infected individuals have some constant level of infectiousness throughout the duration of the incubation interval. Longitudinal studies of the fluctuations in viral abundance or in the concentration of HIV antigens in serum, cells or semen collected from infected patients, however, suggest a more complicated pattern, with two episodes of peak infectivity, one in the early stages of the incubation period and one in the late stages (Pedersen *et al.* 1987; Burger *et al.* 1986; Anderson 1988). In these studies, the concentration of detectable HIV antigen rises to a peak during primary infection, and then typically falls to very low levels (often undetectable by current methods) before beginning to rise again as symptoms of disease appear and the patient progresses from persistent generalized lymphadenopathy (PGL) to ARC and finally to AIDS. Concomitant with these changes in antigenaemia, antibodies to core antigens (anti-p24) rise slowly, reach peak titres when antigen concentration is undetectable or very low, and then fall to very low levels as clinical disease develops. These patterns are illustrated in figure 3.

There are practical problems associated with the quantification of antigen titres in serum samples, further problems in interpretation associated with the relation between antigen concentration in serum and virus abundance in blood, excretions and secretions, and yet further problems in the assumption that infectiousness is proportional to levels of virus in the blood. The patterns in figure 3, however, suggest the tentative hypothesis that there are two phases of infectiousness, one during and immediately after primary HIV infection (of *ca.* 6–12 months duration), and the second as the patient progresses through ARC to AIDS (of duration around 1 year or more), with the two being separated by a period of low infectivity (which may be relatively long if average incubation intervals are around 8 years).

Such time-dependent infectiousness can be represented by defining a conditional transmission coefficient, $\gamma(t, \tau)$, which measures the infectiousness at time t after acquiring infection, for an individual whose incubation interval is τ ($\tau > t > 0$). A rough mathematical representation of the trends described above and in figure 3 is

$$\gamma(t, \tau) = \beta_0 \exp(-t/T_0) + \beta_1 \exp[-(\tau - t)/T_1]. \tag{2.5}$$

Here T_0 and T_1 characterize the duration of the first and second phases of high infectiousness, respectively, and β_0 and β_1 characterize the relative strengths of these two phases (typical parameter values might be $T_0 \approx T_1 \approx 1$ year and $\tau \approx 8$ years (May *et al.* 1988*a*).

The above discussion makes the implicit assumption that all HIV infections cause AIDS. On current evidence, this may not necessarily be so.

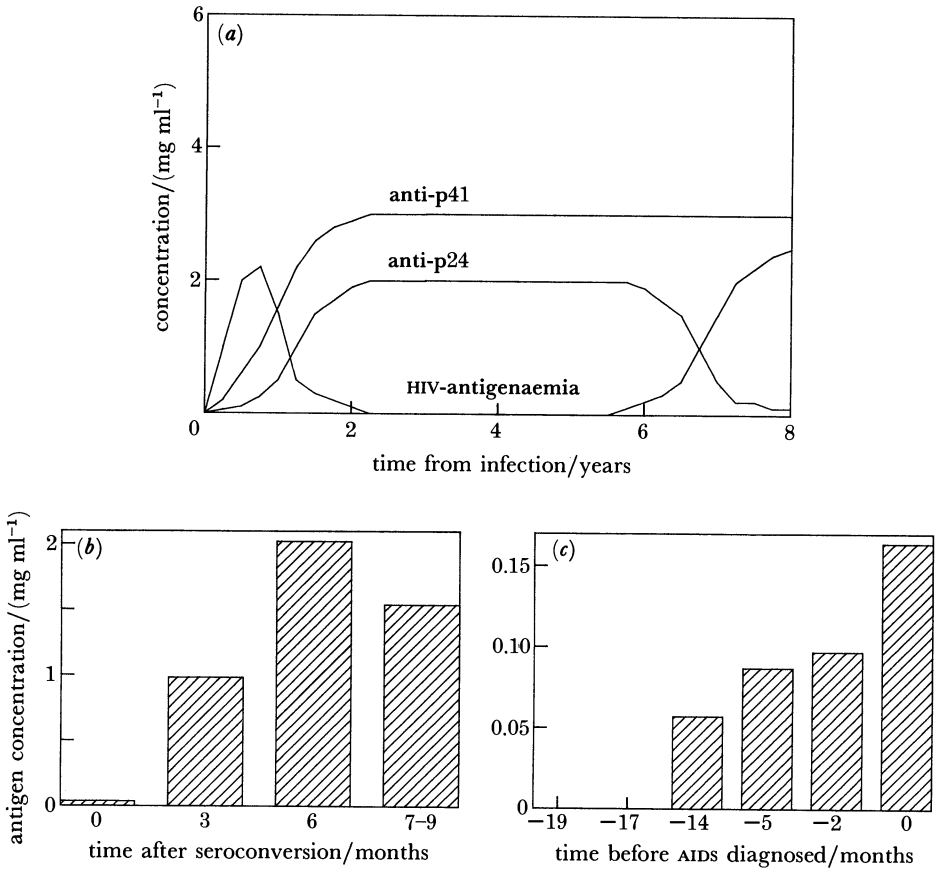

FIGURE 3. (*a*) A schematic representation of changes in antigen and antibody concentrations over the interval from HIV infection to the development of AIDS. The incubation period of AIDS was set at 8 years in this illustration (after Pedersen *et al.* (1987)). (*b*) Average concentrations for antibodies to HIV antigens in serum samples taken from homosexual males infected with HIV, at various times (over the first year) after seroconversion. Sample sizes were: $n = 11$ at time 0, $n = 8$ at 3 months, $n = 4$ at 6 months, and $n = 4$ at 7–9 months (compiled from data surveyed in May *et al.* (1988*a*)). (*c*). Antigen concentrations (in nanograms per millilitre) at specific times, in months, before conversion to AIDS in serum samples drawn from a male homosexual (see May *et al.* (1988*a*)).

Suppose a fraction f do eventually develop AIDS, and the remaining fraction, $1-f$, do not. In simple models with constant infectiousness, the non-AIDS-developing fraction could remain asymptomatic carriers for life (as is arguably the case for asymptomatic carriers of hepatitis B virus, HBV). Alternatively we could assume that this fraction, $1-f$, revert to an uninfectious state at some characteristic rate v' (which may or may not be the same as the rate v of (2.3)). More generally, we could take the relatively realistic expression (2.5) and observe that the non-AIDS-developing fraction, $1-f$, manifest only the first phase of infectiousness (of characteristic duration T_0); for this fraction, effectively $\tau \to \infty$.

The possibility of some fraction's never developing AIDS also affects the normalization of the distribution of incubation times. Equations (2.3) and (2.4) assume the distribution integrates to unity; that is, the probability that an individual will incubate AIDS after some time interval is unity. More generally, the integral over all incubation times will come to f, and the remaining $1-f$ represents individuals who never develop AIDS.

2.4. *Transmission probabilities*

When considering the probability that an infected individual will infect a susceptible sexual partner, one of two assumptions that are opposite extremes may be used. One assumption is that there is, on average, some constant transmission probability, δ, per sexual contact, and that these probabilities compound independently and randomly. The overall probability for an infected individual to infect a sexual partner then increases with the number of contacts, and this Poisson process gives, on average, a transmission probability per partnership of approximately $1-\exp(-\delta n)$ after n contacts. The opposite extreme, which at first sight may seem less likely, is that the transmission probability is, on average, a constant for all partnerships of a given kind.

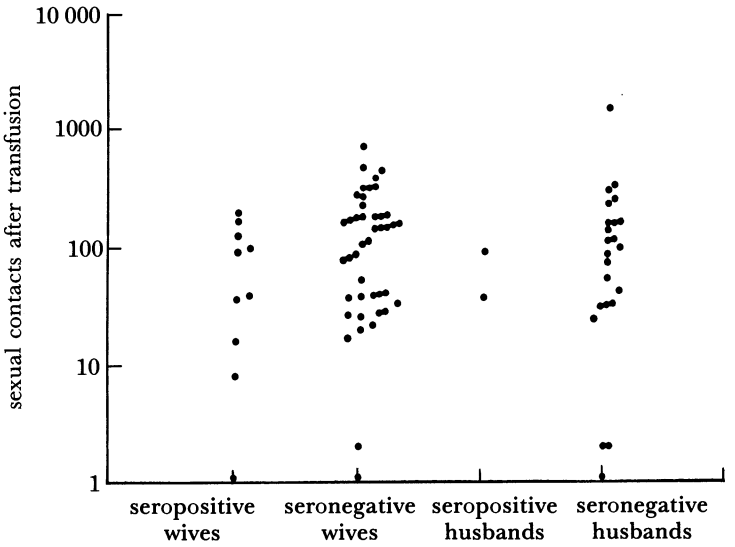

FIGURE 4. For each of 80 individuals, this figure shows the number of sexual contacts with a spouse, after the spouse was (unknowingly) infected with HIV by blood transfusion. The 25 husbands and 55 wives of infected individuals are divided according to current serological status. The figure shows that, in this study group, transmission probability is uncorrelated with number of sexual contacts. (After Peterman *et al.* 1988.)

Figure 4 summarizes relevant data of Peterman *et al.* (1988). From a larger pool of individuals with transfusion-associated AIDS or HIV infection, Peterman *et al.* selected 106 index cases. Of these, 26 had had no sexual contact with their spouse since the transfusion, leaving a pool of 80 infected individuals – 55 men and 25 women – and their spouses. Standard interviews were conducted with the index cases and their spouses, reviewing medical and social histories; the study excluded all individuals who had risk factors other than the infected spouse. In particular, wives/husbands were asked about their sexual practices, about the average number of sexual contacts with their infected husband/wife per month since the transfusion, and about changes in frequency of sexual contacts after the partner had been diagnosed as having HIV seropositivity or AIDS. In all, Peterman *et al.* found that 2 of the 25 husbands of infected wives, and 10 of the 55 wives of infected husbands, were HIV seropositive. Figure 4 shows the relation between transmission of HIV infection and the number of sexual contacts with the infected spouse. The figure points to the surprising conclusion that transmission probability

is unrelated to the actual number of sexual contacts, for partnerships ranging from one to several thousand contacts. Insofar as there is any statistical correlation between transmission probability and number of contacts, it is negative (although not statistically significant): one infected woman had had only a single sexual contact, and another had only 8; 11 of 55 wives and 5 of 25 husbands remained uninfected after more than 200 sexual contacts with their infected spouses. For further discussion, including caveats and speculations about the reasons for the apparent lack of correlation between transmission probability and number of sexual contacts, see Peterman *et al.* (1988) and May (1988).

Given the small sample sizes, the estimate of Peterman *et al.* of an average transmission probability of around 0.1 (2/25) female-to-male and around 0.2 (10/55) male-to-female is consistent with other estimates of transmission probabilities in heterosexual partnerships, summarized in table 2. For many of the studies in table 2, it is by no means clear that infection was acquired by sexual contact with the heterosexual index case. None of these other studies reported data about the correlation between transmission probability and number of sexual contacts.

TABLE 2. HIV SEROPREVALENCE IN HETEROSEXUAL PARTNERS OF INDIVIDUALS INFECTED WITH HIV, FROM STUDIES SUMMARIZED BY HHS (1987)

(Unlike the study by Peterman *et al.* (1988), which is not included here, none of these studies provide information about the correlation between transmission probability and number of sexual contacts.)

country	number of partners tested	sex of partner at risk	fraction seropositive (%)	source of HIV seropositive partners
U.S.A.	21	f	10	haemophiliacs
U.S.A.	19	f	21	haemophiliacs
France	148	f	7	haemophiliacs
U.K.	36	f	8	haemophiliacs
U.S.A.	24	f	17	haemophiliacs
U.K.	14	f	0	haemophiliacs with ARC or AIDS
U.S.A.	21	f	10	haemophiliacs with ARC or AIDS
U.S.A.	4	f	25	transfusion recipients
U.S.A.	55	f	24	bisexual men
U.S.A.	7	f	43	bisexual men
U.S.A.	12	f	42	IV-drug users
U.S.A.	69	f	46	IV-drug users
U.S.A.	5	m	60	IV-drug users
U.S.A.	11	m	55	individuals with AIDS
U.S.A.	45	m and f	58	individuals with AIDS
U.S.A.	42	m and f	48	individuals with ARC or AIDS
U.S.A.	8	f	0	various
U.S.A.	35	f	35	various
U.S.A.	22	m and f	36	various

It is more difficult to use partner tracing to estimate transmission probabilities in homosexual partnerships, because of the relatively high number of multiple partners among homosexual men (Grant *et al.* 1987). Grant *et al.* have used a mathematical model, in conjunction with seroprevalence studies and information about the sex history of individual homosexual men, to estimate that the average transmission probability per partnership (which they call the infectivity), associated with the practice of unprotected receptive anal intercourse, is around 0.09, with a confidence interval from 0.04 to 0.15. Within this group, there is a pronounced

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

correlation between number of partners and seropositivity levels, as suggested by the mathematical model and as discussed in more detail below. The rough equivalence between this transmission probability and those discussed above for heterosexual partnership is surprising. Despite the evidence of figure 4, we think it may be that a higher transmission probability per sexual act among homosexual men could be counterbalanced by the longer average duration of the heterosexual partnerships studied.

A fully accurate mathematical model for the transmission dynamics of HIV by sexual contacts of a given kind would need to keep track of the formation and break-up of all partnerships, taking account of the infection status of each partner and of the possible dependence of transmission probabilities upon the duration of partnerships. This would be a formidable undertaking. Existing models are based on various kinds of approximations.

One approach, pioneered by Dietz (1988), is to take explicit account of the formation and dissolution of partnerships (either heterosexual or homosexual), but to allow an individual to have at most one partnership at any one time. Although this may be a reasonable basis for approximating heterosexual transmission of HIV in a fairly monogamous society, even in this circumstance we think it likely to underestimate the rate at which infection spreads (because individuals who have several concurrent sexual relationships are likely to play a disproportionate role in the dynamics of transmission). Dietz also assumes some average transmission probability per contact, or per unit time, compounding in Poisson fashion. The basic approach can, however, easily be modified to treat the transmission probability as being roughly a constant per partnership, independent of its duration (which figure 4 suggests may be closer to reality), or as having some intermediate dependence on partnership duration (along lines suggested by Hyman & Stanley (1988)).

A more phenomenological approach, which has been used successfully by Hethcote & Yorke (1984) for modelling gonorrhea, is to assume that the probability (per unit time) for a susceptible individual to acquire infection is equal to the number of sexual partners, i, multiplied by the probability of being infected by any one partner, λ. In turn, λ is given by the probability that a partner is infected multiplied by the probability (per unit time) that infection will be transmitted from such a partner. It is clear that this rough approximation compounds the instantaneous transmission probabilities over successive time intervals, paying no attention to possible correlations among the sexual partners in successive time intervals. As such, the approach may be most accurate when there are relatively large numbers of short-lived partnerships (as for highly active homosexual males or, making the necessary changes, needle-sharing IV-drug abusers).

The same mathematical expression for the rate at which susceptibles acquire infection, $i\lambda$, can be given an alternative biological interpretation. Instead of taking i to be the average number of partners and λ the infection probability per partner per unit time, i may be interpreted as the rate at which new sexual partners are acquired (that is, the number of new partners per unit time) and λ as the probability that any one such newly chosen partner will transmit infection (over the duration of the partnership). This interpretation has the virtue of conforming to the data in figure 4, in that transmission probabilities do not compound in Poisson style (tending to saturate to unity in long-lasting partnerships). The concomitant fault in the approach is the mathematical inconsistency of having an instantaneous probability of infection (or probability per unit time) that involves overall transmission probabilities per partnership. To make this approach more accurate, it would be necessary to take account of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the formation and break-up of partnerships, thus moving toward the models of Dietz but allowing for several concurrent partnerships. If, however, infection is acquired within partnerships on timescales less than, or of the order of, other dynamically relevant timescales, then the above approach may be a reliable approximation.

For transmission among IV-drug abusers, high rates of encounter with new individuals seem likely. Models that deal with rates of sharing needles multiplied by the probability that a given needle will cause infection (compounding Poisson-style) are thus likely to be accurate.

These different mathematical approaches can be put in some perspective by considering the basic reproductive rate, R_0, for HIV infections within a specified risk-group. From the definition given above, R_0 is equal to the average number of new partners acquired per unit time by an infected individual multiplied by the average duration of infectiousness (to give the average number of new sexual partners acquired while infectious), all multiplied by the average transmission probability per partnership. This is common sense, and we could thus obtain a rough estimate of R_0 for any particular risk group from the kinds of data shown in figure 4 (for heterosexual transmission) or in the study by Grant *et al.* (1987) (for homosexual males), along with estimates of average rates of acquiring partners and duration of infectiousness. The difficulties arise, however, in giving more rigorous specification to what is meant by 'average' in the above statements. This will emerge in our discussion of the mathematical models below.

2.5. *Rates of acquiring new partners*

There have, until recently, been surprisingly few studies of average rates of acquiring new heterosexual or homosexual partners by men and women. Even less information is available about the distribution in such rates. Reviews of such data as are available for patterns in developed countries are given by May & Anderson (1987), Anderson (1988) and Johnson (1988); what follows is a sketchy summary.

In San Francisco, a study of an unbiased sample of 814 homosexual/bisexual men gave an average of 10.8 partners in 1984 (with 27% having more than 10 partners per year (Winkelstein *et al.* 1987)). A similar study in London gave an average of 10.5 partners per year in 1986, falling to 4.8 in 1987, following education campaigns on television and elsewhere (Anderson & Johnson 1988). As illustrated in figure 5, these and other data show high variability in rates of acquiring new homosexual partners, with the distributions typically having coefficients of variation (ratios of standard deviations to means) exceeding unity. Both surveys about sexual habits and the recorded incidence of other STDs, such as gonorrhea, suggest significant decreases in the average rates at which new homosexual partners are acquired in the U.S.A. and the U.K. These average rates, however, appear still to be significantly higher than the corresponding rates among heterosexuals.

It is probable that bisexual men currently represent an important source of transmission of HIV infection to women. There are, unfortunately, very few studies of the pertinent rates of acquiring female partners by bisexual men. In one recent study of homosexual men, 30% reported themselves to be bisexual, and of these roughly one-third reported more than two female sexual partners over the past year; this seems to us to be a disturbing figure from the point of view of HIV epidemiology (Anderson 1988).

Studies of heterosexual men and women in developed countries consistently show the average rate of acquiring new partners, or the average number of partners per year or per

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

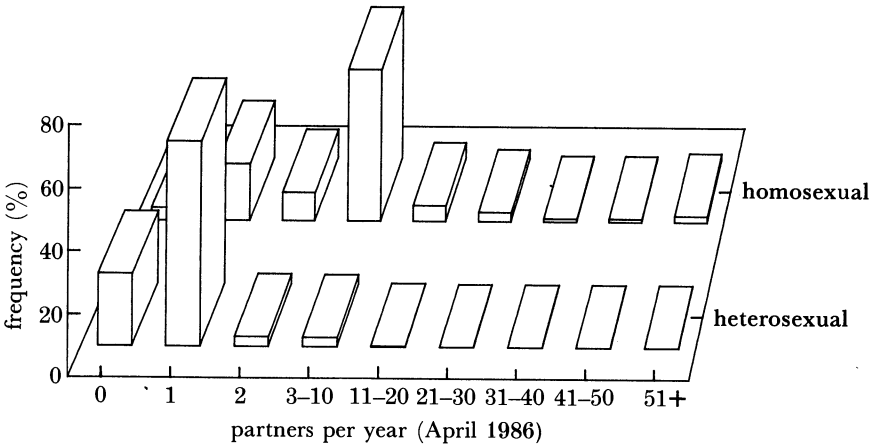

FIGURE 5. Distributions in numbers of sexual partners per year (as of April 1986) among a survey group of homosexual males and among a survey group of heterosexual males and females in Britain (data from BMRB (1987); see also Anderson (1988)). Mean values: homosexuals 8.7, heterosexuals 0.91.

lifetime, to be significantly smaller than for homosexual males. For example, the above-mentioned study by Winkelstein *et al.* (1987) also included 212 heterosexual men, who reported an average of 2.8 female partners in 1984 (with only 2.8% reporting more than 10 partners). In the U.K., several independent surveys of heterosexuals suggest an average of around 1 partner per year for both men and women, and average lifetime totals of around 3–4 for women and around 4–12 for men (Wadsworth *et al.* 1988); see figure 6. Studies of men and women attending STD clinics tend to give higher average values. Again, as illustrated in figures 5 and 6, all these studies tend to exhibit considerable variability in rates of acquiring new partners within any one study group.

Studies that combine information about seropositivity with information about the sexual habits of homosexual males consistently show significant correlations between rates of acquiring new homosexual partners and seropositivity levels (Winkelstein *et al.* 1987; Johnson 1988; May & Anderson 1987). Although less abundantly documented, similar correlations have been found among heterosexuals.

There is very little data about rates of partner change in central African countries. It has been suggested, however, that the patterns of sexual activity of males and females may differ more than is typically the case in developed countries, with the majority of females having relatively monogamous marriages or '*unions libres*' (persistent cohabitation without formal marriage), but where many of the male partners in such relationships are less monogamous, with the books kept in balance by a cadre of young female prostitutes (Quinn *et al.* 1986).

For IV-drug users, the corresponding information would seem to be the typical size of needle-sharing groups, and the typical rate at which individuals enter and leave such groups. A study in Edinburgh, for example, suggests that needles may be shared among 10–20 drug abusers (Robertson *et al.* 1986). The ethnography of needle-sharing, however, differs greatly from place to place. Among some groups there is apparently a practice of drawing blood into the syringe, to mix it with the drug before injecting; the ritual can be to leave some blood in the syringe (a token of sharing, as it were). In this event, the total number sharing the needle can be less important than the HIV status of the previous individual using the needle. These complexities

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

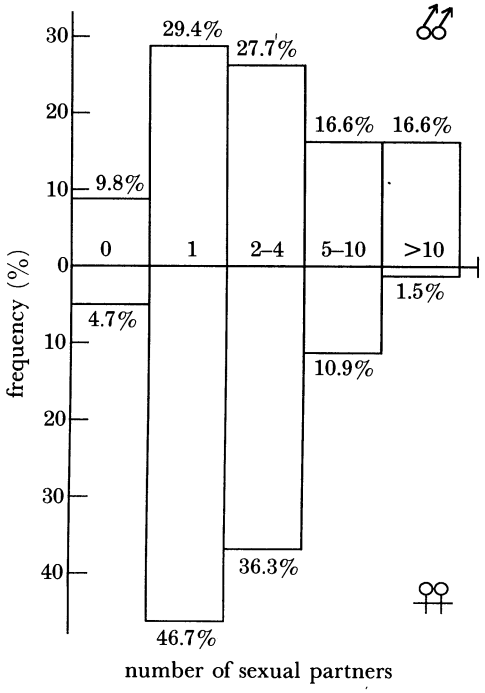

FIGURE 6. Distributions in reported lifetime numbers of sexual partners for a random sample of 296 heterosexual males and 405 heterosexual females, between the ages of 16 and 64, carried out in Britain in early autumn 1987 (Wadsworth *et al.* 1988). Note that males report a significantly higher average number of lifetime partners; this imbalance may derive partly from age-structure effects within the samples or from inaccurate reporting, but it probably arises mainly because very promiscuous females (such as prostitutes) are not represented in this relatively small sample.

matter. One of the most important public health measures could, for instance, be to persuade the managers of the illegal drug trade that it is in their long-term interest to ensure sterile needles in 'shooting galleries'.

3. A BASIC MODEL

We begin by considering a very simple model, to which various kinds of refinement will be added.

Let the total population under consideration be $N(t)$ at time t. This population may be homosexual males in a large city, or the total population of heterosexuals (with male-to-female and female-to-male overall transmission rates assumed equal, so that the model effectively has the structure of a one-sex model), or any other defined group. Let $N(t)$ be subdivided into $X(t)$ susceptibles and $Y(t)$ infecteds, who are assumed also to be infectious. It is further assumed that all infecteds move at a constant rate v (that is, after an average incubation time $1/v$) to develop full-blown AIDS, at which point they are regarded for the purposes of the model as being effectively removed from the group under study (that is, $N(t) = X(t) + Y(t)$); in fact, the average life expectancy of diagnosed AIDS patients is 12.6 months in the U.K. (Reeves & Overton 1988) and 11.4 months in the U.S.A. (Rothenberg *et al.* 1987). Deaths from all other causes occur at a constant rate per head μ, and new susceptibles appear at the overall rate $B(t)$.

This simple system is described by the pair of first-order differential equations (May *et al.* 1988*b*; Anderson *et al.* 1988):

$$dX/dt = B-(c\lambda+\mu)\,X, \tag{3.1}$$

$$dY/dt = c\lambda X-(v+\mu)\,Y. \tag{3.2}$$

The dynamics of the total population, $N = X+Y$, thus obeys

$$dN/dt = B-\mu N-vY. \tag{3.3}$$

Here $c\lambda$ is the 'force of infection', representing the probability per unit time that a given susceptible will become infected. As discussed in §2, c may be taken as the average rate at which new partners are acquired, and λ to be the probability of acquiring infection from any one partner (Hethcote & Yorke 1984). Further, we can write

$$\lambda = \beta Y/N. \tag{3.4}$$

Here β is the probability of acquiring infection from any one infected partner, and Y/N is the probability that a randomly chosen partner will be infected. Finally, the input of new susceptibles is assumed to be proportional to the size of the population:

$$B(t) = \nu N(t). \tag{3.5}$$

This is a natural first approximation if we are dealing with (symmetrical) heterosexual transmission of HIV in the total population, in which case ν is the overall average birth rate per head. More generally, ν is a rate characterizing the entry of new members into the group in question.

The above system of equations now reduces to a pair of equations for $N(t)$ and $Y(t)$, which can be written

$$dY/dt = [\Lambda-\beta c(Y/N)], \tag{3.6}$$

$$dN/dt = [r-v(Y/N)]. \tag{3.7}$$

Here the definitions Λ and r are introduced:

$$\Lambda \equiv \beta c-\mu-v, \tag{3.8}$$

$$r \equiv \nu-\mu. \tag{3.9}$$

As shown below, Λ is the initial exponential growth rate of the infection (from very low values) within the population, and r is the rate at which the population grows in the absence of AIDS.

Equations (3.6) and (3.7) can be solved analytically. The fraction infected at time t is (May *et al.* 1988*b*)

$$\frac{Y(t)}{N(t)} = \frac{\Lambda \exp(at)}{1+(b/a)\,\Lambda[\exp(at)-1]}, \tag{3.10}$$

and the total population size is

$$N(t) = N(0)\,e^{rt}[1+(b/a)\,\Lambda(e^{at}-1)]^{-v/b}. \tag{3.11}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Here Λ is the initial fraction infected, at $t = 0$. The quantities a and b have been defined for notational convenience: $a \equiv \Lambda - r, b \equiv a + \nu$.

Two things are immediately apparent. First, let us neglect demographic parameters in comparison with epidemiological ones by assuming the birth and death rates per head (ν and μ) are significantly less than the transmission and incubation rates (βc and v), so that $a \approx \Lambda \approx \beta c - v$. Equation (3.10) then shows that the fraction of the population infected with HIV initially grows exponentially at rate Λ. These results have an intuitive explanation: infected individuals on average transmit HIV, with probability β, to c new partners each year, while a fraction v of infected individuals go on to develop AIDS; hence the exponential growth rate of infection among the population is $\beta c - v$. That is, the incidence of HIV infection (and thus, very roughly, the subsequent incidence of AIDS) is expected initially to grow exponentially with a doubling time $t_d \approx (\ln 2)/\Lambda$. Table 3 summarizes such doubling times for various risk groups in different countries. In particular, early doubling times for the incidence of infection among male homosexuals in developed countries are consistently around 0.5 to 1 years. That is, $\Lambda \approx 1$ per year for such groups. Finally, if the average incubation time is indeed around 8 years or so, this very simple model, in combination with the data summarized in table 3, suggests that for homosexual males in developed countries the product βc is around 1 per year. This wholly independent estimate, based on population-level data, is consistent with the estimates $\beta \approx 0.1$ and $c \approx 10$ per year for homosexual males, as discussed in §§2.4 and 2.5 respectively.

Table 3. Doubling times based on AIDS case reports to the World Health Organization

country	time period	doubling time/months
Australia	1986–87	12.9
Austria	1983–85	15.6
Bahamas	1986–87	12.7
Brazil	1983–85	8.5
Canada	1983–85	12.7
Dominican Republic	1986–87	6.4
France	1983–85	11.2
Greece	1986–87	7.8
Jamaica	1986–87	5.8
Kenya	1986–87	4.9
Mexico	1986–87	7.5
Netherlands	1983–86	11.2
Portugal	1986–87	11.8
South Africa	1986–87	11.8
Spain	1983–86	6.3
Sweden	1983–86	9.9
Switzerland	1983–86	10.0
Tanzania	1986–87	8.3
United Kingdom	1983–86	14.0
United States	1981–86	13.0
Venezuela	1986–87	11.2
Zambia	1986–87	13.9

Secondly, if the demographic parameters are retained, then (3.10) shows that $\Lambda > r\,(a > 0)$ is required for HIV/AIDS to establish itself; for $r > \Lambda > 0$, HIV infects an increasing number, but a decreasing proportion, of the exponentially growing population. Given that the infection can establish itself ($a > 0$), (3.11) shows it will have a demographic impact, reducing the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

asymptotic rate of population growth from the disease-free rate, r, to some lower rate, ρ, given by

$$\rho = r - v(\Lambda - r)/(\Lambda + \mu). \tag{3.12}$$

Thus HIV/AIDS can even lead to population decline (negative rates of population growth, $\rho < 0$), provided that Λ is large enough ($\Lambda > r(v+\mu)/(v-r)$). A more realistic discussion of the demographic implications of HIV/AIDS is given in §6.

4. HIV EPIDEMICS IN HETEROGENEOUS POPULATIONS

We now extend the basic model to take account of various kinds of heterogeneity within the risk-group under consideration. In much of this analysis we consider the transmission dynamics of HIV in closed groups, with no recruitment of new susceptibles (that is, $B = 0$).

4.1. *Heterogeneity in rates of acquiring sexual partners*

Consider a closed population of homosexual males (or, *mutatis mutandis*, IV-drug abusers, or – under certain symmetry assumptions – heterosexuals), whose magnitude is $N(t)$ at time t. This population is divided into sub-groups, N_i, whose members on average acquire i new sexual partners per unit time. Initially $N_i(0) = N(0)\,p(i)$, where $p(i)$ is the initial probability distribution in rates of acquiring partners. The number of susceptible, infected (and infectious), and no longer infectious individuals in the ith class are defined to be X_i, Y_i and Z_i respectively, so that $X_i + Y_i + Z_i = N_i$ (it may, of course, be that $Z_i = 0$). If deaths from causes other than AIDS are ignored in this closed population (that is, $\mu = 0$), then (3.1)–(3.3) are replaced by

$$\mathrm{d}X_i/\mathrm{d}t = -i\lambda X_i, \tag{4.1}$$

$$\mathrm{d}Y_i/\mathrm{d}t = i\lambda X_i - vY_i, \tag{4.2}$$

$$\mathrm{d}N_i/\mathrm{d}t = -fvY_i. \tag{4.3}$$

Here we have assumed that a fraction, f, of those infected with HIV go on to develop AIDS – whereupon they are effectively removed from the population under consideration – at a constant rate, v (corresponding to an average incubation time of $D = 1/v$). The remaining fraction, $1-f$, are assumed to become uninfectious at the same rate. As discussed more fully in §2.3, the realities are more complicated. In particular, it may be that those who do not develop AIDS remain infectious indefinitely, or they may be infectious only for the initial phase shown in figure 3; in the absence of detailed knowledge, the preliminary assumption that such individuals lose infectiousness at the same rate as those who develop AIDS seems appropriate.

The infection probability per partner, λ, is now given by generalizing (3.4)

$$\lambda = \beta \frac{\sum iY_i}{\sum iN_i}. \tag{4.4}$$

In (3.4) the probability that any one partner is infectious is simply Y/N; in (4.4) partners are weighted according to their degree of sexual activity, i. By assuming that partners are chosen randomly (apart from the activity levels characterized by the weighting factor i), we may be overestimating the contacts of less active individuals with those in more active categories, and thus overestimating the spread of infection among such less active sub-groups. Conversely, the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

transmission probability β may be higher for longer-lasting partnerships (despite the data in figure 4), so that use of a constant β may tend to underestimate the spread of infection among less active people. The net effect of these countervailing refinements is hard to guess.

Studies of the dynamical behaviour of the system of equations (4.1)–(4.4) is facilitated by defining the quantity $\phi(t)$:

$$\phi(t) = \int_0^t \lambda(s)\,\mathrm{d}s. \tag{4.5}$$

If the initial seed of infection is taken to involve a negligible number of individuals, the initial value of X_i is $X_i(0) = N_i(0) = Np(i)$, and (4.1) for $X_i(t)$ can be integrated to give

$$X_i(t) = Np(i)\,\mathrm{e}^{-i\phi}. \tag{4.6}$$

The factor i in the exponent means that susceptibility is depleted faster in the more highly active groups, as intuition would suggest. Substitution from (4.3) for Y_i in (4.4) for $\lambda(t)$, and integration over time, gives the useful result

$$\sum iN_i(t) = N\langle i\rangle \exp(-fv\phi/\beta). \tag{4.7}$$

Here we have defined $\langle i\rangle$ to represent the expectation value of i (the initial mean rate of acquiring partners) over the distribution $p(i)$; more generally, we define

$$\langle F(i)\rangle \equiv \sum p(i)\,F(i). \tag{4.8}$$

Finally, we can add (4.1) and (4.2), integrate over time, substitute the resulting expression for Y_i in (4.4) for λ, and then use (4.6) and (4.7) to arrive at a first-order differential equation for the dynamical variable $\phi(t)$:

$$(\mathrm{d}\phi/\mathrm{d}t)\exp(-fv\phi/\beta) = \frac{\beta\langle i[1-\exp(-i\phi)]\rangle}{\langle i\rangle} - \frac{\beta}{f}[1-\exp(-fv\phi/\beta)] + \lambda(0). \tag{4.9}$$

Here $\lambda(0)$ is the initial value of the infection probability, calculated from (4.4) with the very small numbers of infectious 'seeds', $Y_i(0)$. From the definition (4.5), ϕ has the initial value $\phi(0) = 0$. A more detailed derivation of this and other results is given by May & Jose (1988).

Once the epidemiological parameters $\beta, v, f, \lambda(0)$ and the initial distribution $\{p(i)\}$ are specified, $\phi(t)$ can be calculated, and thence any other epidemiological variable can be evaluated. In particular, the overall fraction of the original population to have experienced infection by time t, $I(t) = 1 - X(t)/N$, is given from (4.6) as

$$I(t) = \langle 1 - \mathrm{e}^{-i\phi}\rangle. \tag{4.10}$$

The cumulative number of AIDS cases up to time t, $C(t)$, is by definition given by $C(t) = fv\sum\int_0^t Y_i(s)\,\mathrm{d}s$. By using (4.2), (4.6) and (4.10), it can be seen that $C(t)$ may be derived from $I(t)$ via the differential equation

$$\mathrm{d}C(t)/\mathrm{d}t + vC(t) = fvNI(t), \tag{4.11}$$

with the initial condition of course being $C(0) = 0$.

Before presenting some numerical examples and discussing their general properties, we consider some limiting cases and the biological insights they provide.

4.1.1. *Early stages of the epidemic*

In the early stages of the epidemic, $\lambda(t)$ and consequently $\phi(t)$ will be small. Differentiating (4.9) in this limit (remembering $\lambda = \mathrm{d}\phi/\mathrm{d}t$, from (4.5)), gives

$$\mathrm{d}\lambda/\mathrm{d}t = \Lambda\lambda + (\mathcal{O}\lambda^2). \tag{4.12}$$

Here the early rate of exponential growth, Λ, is defined as

$$\Lambda = \beta\langle i^2\rangle/\langle i\rangle - v. \tag{4.13}$$

This result can also be obtained directly from (4.2) and (4.4) by putting $X_i \approx N_i$ for the early phases of the epidemic.

Note that (4.13) conforms with the earlier definition of Λ, (3.8), provided that we define c as

$$c \equiv \langle i^2\rangle/\langle i\rangle. \tag{4.14}$$

That is, for epidemiological purposes, the effective value of the average number of new partners per unit time is not the mean of the distribution, but rather is the ratio of the mean-square to the mean. This result simply reflects the disproportionate role played by individuals in the more active groups, who are both more likely to acquire infection and more likely to spread it. Equation (4.14) can alternatively be written

$$c = m + \sigma^2/m, \tag{4.15}$$

where m is the mean and σ^2 the variance of the distribution $\{p(i)\}$. Thus the effective value of the average number of sexual partners for epidemiological purposes can be significantly larger than the simple mean, if the variance is high. To put it another way, efforts directed toward reducing the rate of partner change among the highly active groups are likely to be disproportionately effective in reducing transmission (Anderson *et al.* 1986; May & Anderson 1987).

4.1.2. *Asymptotic fraction ever infected*

For this closed population, the value of ϕ in the limit $t \to \infty$ can be found by putting $\mathrm{d}\phi/\mathrm{d}t = 0$ in (4.9). Ignoring $\lambda(0)$ (the very small initial level of infection that starts the epidemic), the asymptotic value of $\phi, \alpha \equiv \phi(\infty)$, can be obtained from the transcendental equation

$$\alpha = -(\beta/fv)\ln\{1 - f\langle i(1 - \mathrm{e}^{-i\alpha})\rangle/\langle i\rangle\}. \tag{4.16}$$

With α thus determined, it is a simple matter to calculate the fraction ever infected, $I(\infty)$, from (4.10), and other such quantities.

Equation (4.16) has a non-trivial solution only if the quantity R_0 exceeds unity, where R_0 is defined as

$$R_0 \equiv \beta c/v. \tag{4.17}$$

Here c is given by (4.14), and R_0 is the basic reproductive rate for HIV, as discussed in more intuitive terms in §2.

In the limit of a non-lethal infection ($f \to 0$) in a homogeneous population (where on average all individuals have the same number of partners, c), (4.10) and (4.16) reduce to the

Kermack–McKendrick (1927) result, $I = 1 - \exp(-R_0 I)$. More generally, however, for any specified values of f and R_0, the asymptotic fraction ever infected decreases as the heterogeneity in degrees of sexual activity within the population increases (that is, as the coefficient of variation, CV, of the partner-change distribution increases). This is essentially because, other things being equal, the epidemic tends to burn itself out among those in the highly active classes, thus driving the effective value of the reproductive rate of HIV below unity, before a large fraction of those in the low-activity classes have been infected. The greater the variability in degrees of sexual activity, the more pronounced this effect will be.

Figure 7 shows the asymptotic fraction of the population ever to experience infection, $I(\infty)$, as a function of R_0, assuming 50% of those infected go on to develop AIDS ($f = 0.5$). The initial distribution in rates of acquiring new partners is assumed to be a gamma distribution, and the different curves show $I(\infty)$ against R_0 for specified values of the distribution's CV (for details, see Appendix A).

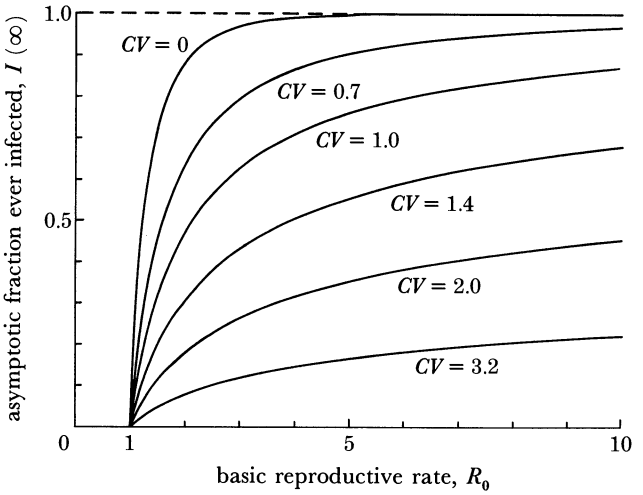

FIGURE 7. Within a closed population of homosexual males, the fraction ever infected during an HIV epidemic, $I(\infty)$, is shown as a function of the basic reproductive rate, R_0, of the infection. The distribution in rates of acquiring new sexual partners within the population is taken to be a gamma distribution, with the coefficient of variation, $CV = \sigma/m$, having the values 0 (classic Kermack–McKendrick epidemic), 0.7, 1, 1.4, 2 and 3.2, as shown. It is assumed that 50% of those infected with HIV eventually die from AIDS (for further details, see text and May (1987)).

Figure 7 bears out the comments made above about the relation between heterogeneity within the population and the asymptotic level of infection for specified R_0. Such a figure can, moreover, be used to make an indirect inference about lower bounds to the value of R_0. As reported in §2, around 50% of homosexual men in some large cities in the U.S.A. are already HIV seropositive. Survey data shows significant heterogeneity in rates of partner acquisition in the early 1980s, corresponding to coefficients of variation of at least unity and probably more. These two observations are hard to reconcile with figure 7 unless R_0 for transmission among such populations is *ca.* 5 or more. Insofar as sexual habits have changed, making for lower average rates of partner acquisition in these populations, we infer higher values for R_0 in the early stages of this epidemic.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

4.1.3. *Dynamics of the epidemic*

From (4.9) with a gamma distribution for the initial distribution $\{p(i)\}$, numerical results can be obtained for the fraction seropositive, the incidence of new AIDS cases, cumulative AIDS cases and so on, as functions of time.

One such result is illustrated in figure 8. The fraction seropositive at first increases exponentially, but this very soon gives way to a more gradual pattern of increase as the highly active classes begin to saturate and new infections come more from the slower dissemination of infection to less active classes. This pattern is significantly different from the classic epidemic of, for instance, measles (where the exponential phase lasts longer), but it accords with the kinds of data discussed in §2 (May & Anderson 1987).

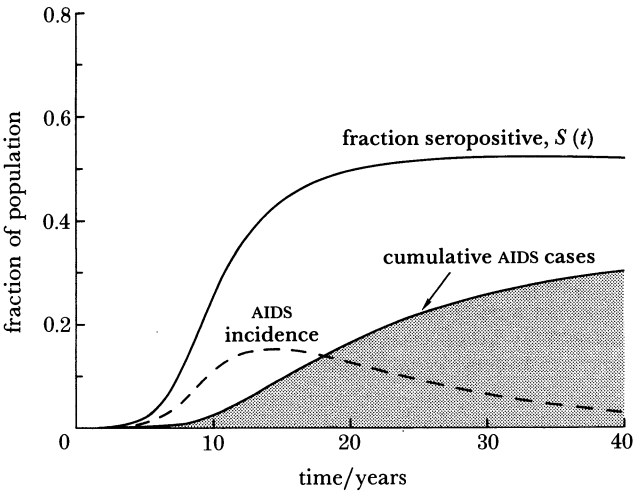

FIGURE 8. This figure shows the fraction seropositive ($S(t) = [NI(t) - C(t)]/[N - C(t)]$), the cumulative number of AIDS cases (as a fraction of the original population, $C(t)/N$), and the rate at which new AIDS cases appear (dC/dt divided by N, and multiplied by 10), as functions of time. These illustrative curves are for a closed population of homosexual males, as described by the mathematical model defined in §4.1 and with a gamma distribution for $\{p(i)\}$; the parameters are $R_0 = 10, v = 0.1$ per year (so that $\beta c = 1$ per year), $f = 0.5$ and $CV = 1.4$ ($\nu = 0.5$) for the sexual-activity distribution. The initial 'seed' of infection is $c\lambda(0) = 0.001$ per year.

These features are made more explicit in figure 9, which shows the progress of the epidemic among separate groups, differentiated according to rates of acquiring new partners. It can be seen that prevalence levels peak relatively soon for the more sexually active categories, and significantly later among the less active categories (May 1988; Anderson *et al.* 1987).

4.2. *Other kinds of heterogeneity*

In seeking to further realism, suppose that – in addition to the heterogeneity in sexual habits discussed above – individuals may also be subdivided into categories, labelled by an index k ($k = 1, 2, \ldots, n$), that differ in their reaction to exposure to HIV infection. These categories may be associated with 'cofactors' such as other STDs or intrinsic genetic differences. If we assume that these categories are not correlated with sexual activity levels (which, as discussed further in §4.5, would not be true with the categories derived from other STDs), we may write $N_{i,k}$ to denote the number of individuals in the ith sexual-activity class and kth

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

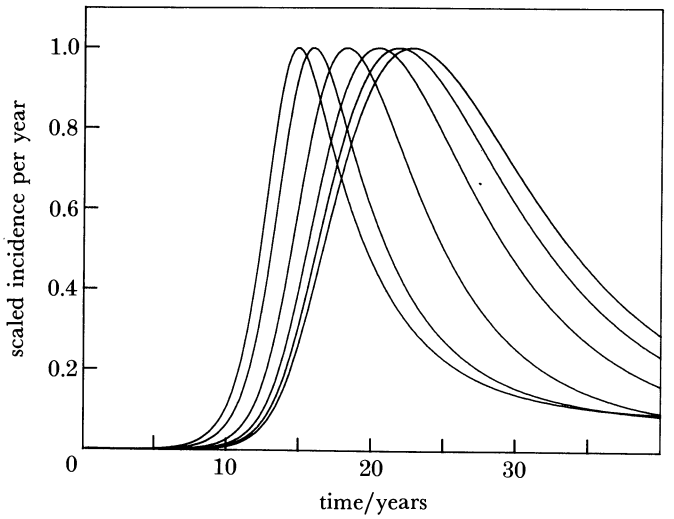

FIGURE 9. This figure shows HIV incidence, or rate at which new HIV infections appear, as a function of time, for each of 6 categories of sexual activity within a population of homosexual males. From right to left, the 6 curves are for the subgroups with 0–1, 2–5, 6–10, 11–50, 50–100 and > 100 new sexual partners per year, respectively. The underlying mathematical model is as defined in §4.1, with a gamma distribution for $\{p(i)\}$; the parameters here are $R_0 = 5, v = 0.2$ per year (so that $\beta c = 1$ per year), $f = 0.3$, and $CV = 1.4$ ($\nu = 0.5$). For each of the 6 groups, the HIV incidence curve has been scaled to have magnitude 1 at its peak, in order to facilitate comparison among the subgroups (on an absolute scale, the curves for the high-activity groups would be scarcely discernible, because these groups include relatively few individuals). It can be seen that the incidence of HIV infection peaks much sooner among the groups with high rates of acquiring new sexual partners, with peak incidence in the lowest-activity group occurring a decade later than the peak for the highest-activity group. Fur further discussion, see Anderson *et al.* (1986, 1987).

epidemiological category; initially the proportion of the population in the kth category is $q(k)$, so that $N_{i,k}(0) = Np(i)\,q(k)$. $N_{i,k}$ may be further partitioned into susceptible, infected, and possibly no-longer-infectious groups, $X_{i,k}$, $Y_{i,k}$ and $Z_{i,k}$, as before.

The dynamical behaviour of $X_{i,k}$, $Y_{i,k}$ and $N_{i,k}$ will again be given by (4.1)–(4.3), with the difference that the rate of leaving the infectious class is v_k (average incubation time $1/v_k$), the fraction going on to develop AIDS is f_k, and the infection probability is λ_k, each of which may be specific to members of the kth category. Equation (4.4) for λ becomes

$$\lambda_k = \frac{\sum_l \beta_{kl} \sum_i i Y_{i,l}}{\sum_i i N_i}. \tag{4.18}$$

Here β_{kl} is the transmission probability for a partnership in which the infected partner is in the lth category and the susceptible in the kth.

For any given set of assumptions about the epidemiological parameters v_k, f_k, β_{kl}, and the initial sexual-activity distribution $\{p(i)\}$, this set of equations can be solved numerically.

Some general insight can, however, be obtained by considering the early phase of the epidemic where, as discussed in §4.1.1, we may put $X_{i,k} \approx N_{i,k}$. The dynamical variables in the ensuing set of linear differential equations may now have their time-dependence factored out as $\exp(\Lambda t)$, with Λ given by

$$\sum_l \left\{ \frac{c\beta_{kl}\, q(l)}{\Lambda + v_l} - \delta_{kl} \right\} \lambda_l = 0. \tag{4.19}$$

Here c is as defined by (4.14); for details of the derivation, see May & Jose (1988). The terms in the curly brackets represent the elements of an $n \times n$ matrix, whose determinant must vanish if (4.19) is to be satisfied. A more explicit result for Λ can be obtained by making the reasonable assumption that the transmission probability can be written as a product, $\beta_{kl} = g_k h_l$, where h_l characterizes the infectiousness of individuals in the lth category, and g_k the susceptibility in the kth. In this event, the requirement that the determinant of the matrix defined by (4.19) should vanish gives

$$\sum \frac{c\beta_{kk}\, q(k)}{(\Lambda + v_k)} = 1. \tag{4.20}$$

Equation (4.20) will give growth rates, Λ, with positive real parts only if $R_0 > 1$, where the definition (4.17) for R_0 is generalized to

$$R_0 \equiv \sum c\beta_{kk}\, q(k)/v_k. \tag{4.21}$$

For a derivation and further discussion of this result, see May & Jose (1988) and May (1987).

Equation (4.21) may be used to illustrate the connection between f, the fraction of HIV infecteds who go on to develop AIDS and die from it, and the total number dying in an HIV epidemic in a closed population. This connection can be somewhat paradoxical, in that increasing f does not necessarily mean a larger total of deaths. To see this, let us assume a fraction f of those infected develop AIDS after an average incubation time $1/v_1$, whereas the remaining fraction, $1-f$, remain infectious for a characteristic time $1/v_2$. Both groups are assumed to be equally infectious, with transmission probability β; this assumption can easily be generalized. We assume, however, that $1/v_1$ is significantly smaller than $1/v_2$ (with these times possibly being something like 8 and 30 years, respectively), so that v_1/v_2 significantly exceeds unity. Then R_0 has the form

$$R_0 = [c\beta/v_1]\,[f + (1-f)\,(v_1/v_2)]. \tag{4.22}$$

Thus R_0 can be significantly larger for smaller values of f than for larger ones, by virtue of the factor v_1/v_2. The overall fraction of the original population experiencing infection, $I(\infty)$, is larger for larger values of R_0. The total number of AIDS deaths is proportional to f multiplied by $I(\infty)$; the first of these two factors obviously increases with increasing f, but the second decreases with increasing f. Thus the total number of AIDS deaths does not bear a simple relation to f in this case. Numerical studies suggest the total number of AIDS deaths initially increases roughly linearly with increasing f, but remains roughly independent of f for f-values above 50% or so; if v_1/v_2 is large enough, total deaths might even decline somewhat for f-values approaching 100% (May & Anderson 1987; May & Jose 1988). All this makes sense: if a substantial proportion of infected individuals remain asymptomatic carriers effectively for life, more infections will be produced. This phenomenon, of course, depends on the existence of long-lived asymptomatic carriers. Current uncertainties about these matters are a major obstacle to determining whether R_0 is likely to exceed unity for heterosexual transmission in developed countries.

4.3. *Distributed incubation times and variable infectiousness*

Returning to §2.3, we recall that the rate of progression from HIV infection to AIDS, $v(\tau)$, is not a constant (as our models have assumed up to this point) but rather depends on the time,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

τ, since infection was acquired. In addition, transmission probabilities may depend on τ, with the conditional transmission probability for an individual whose incubation time is s being $\beta(\tau; s)$ at time τ since infection.

This makes for substantial complications in the analysis, some of which are now sketched. For a fuller discussion of the biological implications see May *et al.* (1988*a*) and Anderson (1988), and for mathematical details see May & Josė (1988) and Castillo-Chavez *et al.* (1988).

To begin, let the index k of §4.2 label individuals according to the duration of their incubation times, with individuals in the kth category having incubation times of length k; this corresponds to putting $v_k(\tau) = 0$ for $\tau < k$ and $v_k(\tau) = \infty$ for $\tau > k$, for individuals in the kth category. We now define $Y_{i,k}(t, \tau)$ to be the number of individuals who have been infected for a time interval τ (at time t), and who are in the ith sexual-activity class and have incubation times of duration k (Anderson *et al.* 1986). This quantity obeys the partial differential equation obtained by generalizing (4.2),

$$\partial Y_{i,k}/\partial t + \partial Y_{i,k}/\partial \tau = -v_k(\tau)\, Y_{i,k}(t, \tau). \tag{4.23}$$

One boundary condition is given by the rate at which new susceptibles appear,

$$Y_{i,k}(t, 0) = i\lambda X_{i,k}(t). \tag{4.24}$$

The other boundary condition specifies $Y_{i,k}$ for all τ at some initial time $t = 0$. The infection probability λ obeys the appropriate generalization of (4.18):

$$\lambda = \frac{\sum_k \sum_i i \int_0^k \beta(\tau; k)\, Y_{i,k}(t, \tau)\, \mathrm{d}\tau}{\sum i N_i}. \tag{4.25}$$

We have assumed that all individuals are equally susceptible.

Once the functional form of $\beta(\tau; k)$ and the distribution of incubation times are specified, the dynamical system defined above can be studied numerically for any chosen set of epidemiological parameters.

Again, however, some general insights can be gained by considering the early phase of the epidemic in which $X_{i,k} \approx N_{i,k} = Np(i)\, q(k)$. As before, time dependences can be factored out as $\exp(\Lambda t)$, and (4.23) can then be integrated to obtain

$$Y_{i,k}(\tau) = i\lambda N p(i)\, q(k)\, \mathrm{e}^{-\Lambda\tau}, \tag{4.26}$$

for $\tau < k$; $Y_{i,k}$ is zero for $\tau > k$. A more rigorous solution of (4.23), using Laplace transform techniques, is presented elsewhere (May & Jose 1988). Equation (4.26) may now be substituted into (4.25), to give a 'dispersion relation' for Λ:

$$1 = c \sum_k q(k) \int_0^k \beta(\tau; k)\, \mathrm{e}^{-\Lambda\tau}\, \mathrm{d}\tau. \tag{4.27}$$

Finally, as explained in §2.3, the probability of developing AIDS after an incubation interval of duration k, $q(k)$, may be expressed in terms of the time-dependent rate process, $v(k)$, as $q(k) = v(k) \exp\left[-\int_0^k v(s)\, \mathrm{d}s\right]$. Substituting this into (4.27), the early exponential growth rate, Λ, is given by

$$1 = c \int_0^\infty v(k) \exp\left[-\int_0^k v(s)\, \mathrm{d}s\right] \int_0^k \beta(\tau; k)\, \mathrm{e}^{-\Lambda\tau}\, \mathrm{d}\tau\, \mathrm{d}k. \tag{4.28}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

If $\beta(\tau)$ has no conditional dependence on the incubation interval (which is unlikely), (4.28) can be brought into simpler form

$$1 = \int_0^\infty c\beta(\tau) \exp\left[-\Lambda\tau - \int_0^\tau v(s)\,\mathrm{d}s\right]\mathrm{d}\tau. \tag{4.29}$$

This is the so-called Euler equation of mathematical demography. This result has a sensible interpretation, with $c\beta(\tau)$ being the 'fecundity' of HIV infections of 'age' τ, $\exp[-\int v\,\mathrm{d}s]$ the age-specific survivorship function, and Λ the rate of growth of the HIV 'population'. More generally, however, the fecundity of a given individual is at all times contingent upon the life expectancy of that individual, as shown in (4.28).

In §2.3 we discussed the possibility that there may typically be two phases of peak infectivity, separated by a relatively long episode of lower infectiousness. This can have significant implications for the interpretation of temporal trends in the incidence of AIDS and for the estimation of epidemiological parameters. Suppose, for example, $\beta(\tau;k)$ is given by the phenomenological equation (2.5) discussed earlier; if the two episodes of infectiousness are short in relation to the total incubation period, and if the doubling time of the epidemic is around 1 year or so, then to a good approximation the early growth rate is given by $\Lambda \approx c\beta_0 - 1/T_0$. But the basic reproductive rate involves both episodes of peak infectiousness, $R_0 \approx c(\beta_0 T_0 + \beta_1 T_1)$. This is in marked contrast with the simpler models with constant infectiousness, where comparison of (4.13) and (4.17) show Λ and R_0 to have a very direct relation to each other, $\Lambda = v(R_0 - 1)$. The above observation that Λ may often depend mainly on the parameters characterizing the first phase of infectiousness, whereas R_0 involves both phases, is intuitively reasonable: in the terminology of conventional demography, early 'births' count more than later ones toward population growth rates, but all 'births' are relevant to the total number of offspring (here meaning infected people).

More thorough consideration of these complications (May *et al.* 1988*a*) suggests that if indeed there are two peak phases of infectiousness that are both significantly shorter than the average incubation period for AIDS, then current estimates of the transmission coefficients (based on knowledge of the doubling time of the epidemic and the assumption that individuals are infectious over the entire incubation period of 8 years or more) are likely to be significant underestimates of the true likelihood of transmission during any given infectious episode. Such fluctuations in infectiousness, moreover, can induce complex temporal patterns in the epidemic curves, which make it harder to use the models to analyse data and to make predictions.

4.4. *Changes in sexual habits over time*

All the above models assume that, on average, individuals do not change their levels of sexual activity over time. At least for homosexual males this is transparently not the case, as discussed in §2.5.

The basic model of §3, which dealt simply with an average rate of acquiring new partners, c, can easily be modified to take account of changing sexual patterns, by letting $c(t)$ be time-dependent. Such changes may be deduced from survey data or inferred from data about other STDs. Numerical calculations can then be done for any specific assumption about how $c(t)$ changes over time.

For the basic model defined by (3.6) and (3.7), which combines epidemiology with

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

demography, it happens that the analytic expression (3.10) for the fraction seropositive at time t can be generalized for time-dependent $c(t)$ to give

$$\frac{Y(t)}{N(t)} = \frac{\Delta \exp[A(t)]}{1+\Delta\{\exp[A(t)]-1+\nu\int_0^t \exp[A(s)]\,ds\}}. \tag{4.30}$$

Here Δ is again the initial fraction infected (who start the epidemic), and $A(t)$ is defined as $A(t) = \int_0^t [\beta c(s) - v - \nu]\,ds$; $A(t)$ reduces to the at of (3.10) if c is a constant.

Figure 10 compares the pattern of seropositivity over time given by the basic model with constant c, with that given by (4.30) for a population where the average rate of acquiring new partners decreases as $c(t) = c(t_0)/[1+(t-t_0)/T]$ for times $t > t_0$; here t_0 is the time when seropositivity first exceeds 5%. Figure 10 shows results for several values of the parameter T, which measures the average time taken for $c(t)$ to halve its original value. Remember, $f = 1$ in these models, so that Y/N measures the fraction seropositive. It is apparent that such systematic reductions in levels of sexual activity can produce epidemic curves that depart from the classic, measles-like, exponential growth pattern to show early deceleration.

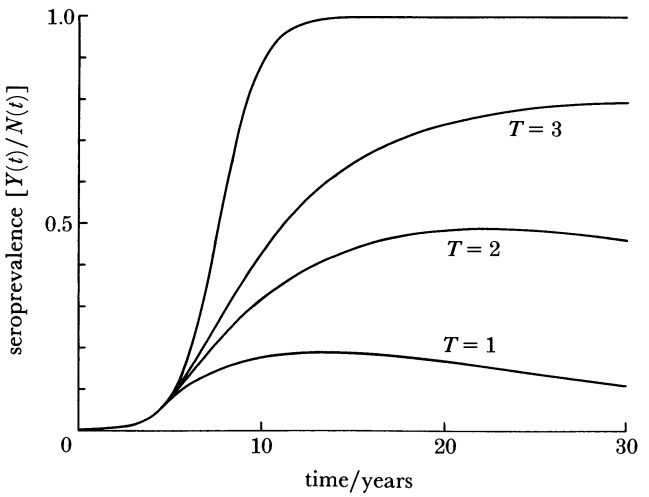

FIGURE 10. The pattern of seropositivity, $S(t)$, over time in a closed population of homosexual males for the basic model of §3 is contrasted with the patterns that arise when average levels of sexual activity change over time (as discussed in §4.4); in these models, $S(t) = Y(t)/N(t)$. Specifically, this figure compares the basic model with an unchanging rate of partner change, c (the top curve), with models in which numbers of partners decrease as $c(t) = c(t_0)/[1+(t-t_0)/T]$ for $t > t_0$, where t_0 is the time at which seropositivity first exceeds 5%. As shown, the curves are for $T = 1$, 2 and 3 years; the other epidemiological parameters in (3.10) and (4.30) are $\beta c = 1$ per year, $v = 0.1$ per year, $\nu = \mu = 0$, and the 'seed' of infection is $\Delta = 0.001$.

In figures 8 and 9 we noted that observed trends in HIV seropositivity appear to move from early phases of roughly exponential growth into slower (and more nearly linear) growth phases, and we indicated how this is to be expected if there is significant heterogeneity in levels of sexual activity. In these heterogeneous models the more rapid spread of HIV, and thence AIDS, among more sexually active groups results in average rates of partner change decreasing over time (as more active individuals are removed), even though individuals are not assumed to change their activity levels. Indeed, an explicit expression can be obtained for the change in the mean

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

number of new partners per unit time, $m(t)$, for the models of heterogeneous transmission in a closed population that were discussed in §4.1:

$$m(t) \equiv \frac{\sum i N_i(t)}{\sum N_i(t)}. \tag{4.31}$$

The numerator on the r.h.s. is given in terms of the dynamical variable $\phi(t)$ by (4.7), and the denominator can be expressed as $N-C(t)$, where N is the original population size and $C(t)$ is the cumulative number of AIDS cases, given by (4.11). Assuming a gamma distribution for the initial $\{p(i)\}$, the overall change in average numbers of partners can be calculated, over time, by using the same epidemiological parameters as in figures 7–9. The results are shown in figure 11, and it can be seen that the dynamics of the epidemic can – by differentially removing more active individuals – produce a marked decrease in overall levels of sexual activity, even though individuals do not change.

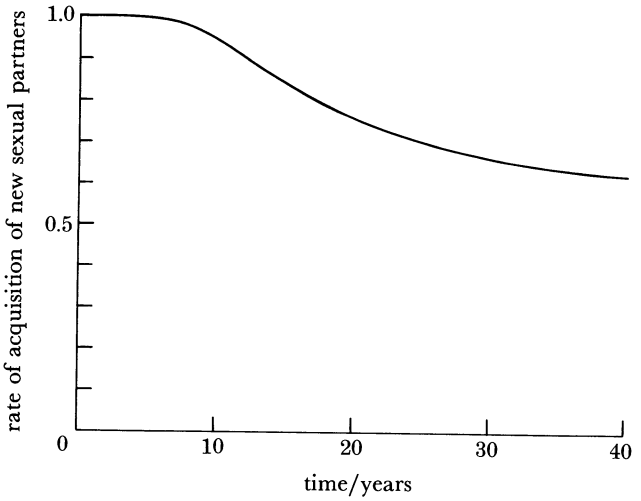

FIGURE 11. This figure shows the change in the average rate at which new sexual partners are acquired (expressed as a ratio to the initial rate, $m(t)/m(0)$), as a function of time, t. Even though no individuals are assumed to change their sexual habits, this change in the overall average comes about because individuals who are more sexually active are more likely to acquire HIV infection, and thus to be removed from the population by dying from AIDS. This illustrative example is based on the mathematical model discussed in §4.1, with a gamma distribution for $\{p(i)\}$; the parameters are $R_0 = 10, v = 0.1$ per year, $f = 0.5$ and $CV = 3.2$ ($\nu = 0.1$).

In other words, the epidemic curves in figure 10 are based on a homogeneous model, in which the average rates of partner change of individuals decrease over time in response to social changes. In contrast, figures 8 and 9 are based on heterogeneous models, in which individuals do not change their sexual habits; here, overall average rates of partner-change fall as a result of the dynamics of the epidemic. In reality, both processes operate. Although we believe that changes in patterns of sexual activity among populations of homosexual men come predominately from changes in individual behaviour, the additional changes produced by the epidemiological dynamics may not be negligible, and deserve to be recognized.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

4.5. *HIV infection in conjunction with other STDs*

It is increasingly believed that heterosexual transmission of HIV in Africa and elsewhere may be facilitated by the presence of other STDs, which can cause sores or ulcers, or otherwise help infectious agents to penetrate the skin. Insofar as this is the case, such other STDs could be expected to be differentially present among those with high rates of partner change, further enhancing such individuals' disproportionate role in acquiring and transmitting HIV infection. There is the further implication that intervention programmes aimed at reducing the incidence of these other STDs could retard the spread of HIV. A clearer understanding of the interplay between the transmission dynamics of HIV and other STDs can help us evaluate the likely efficacy of such a programme. (In passing, we note that a high prevalence of HIV infection among those with other STDs does not necessarily mean that such STDs indeed facilitate HIV transmission; this could easily be correlation without causation.)

We now sketch the essential conclusions of such an analysis. The details will be presented elsewhere. Suppose that, as in §4.1, the population at risk is divided into sub-groups, N_i, in which an average of i new sexual partners are acquired per unit time. The endemic prevalence of a given STD (or some aggregation of STDs) can now be estimated; this prevalence will, of course, be higher among the more sexually active groups. Now HIV is introduced. As a simplest first approximation, we assume that the transmission probability for HIV, β, is enhanced by the factor a ($a > 1$) if either the infectious or the susceptible partner has another STD, and by a^2 if both do. The dynamics of this system can now be studied for any given set of epidemiological parameters.

As ever, it is easier to obtain analytic results for the early stages of the epidemic, when the equations can be linearized and time-dependences factored out as $\exp(\Lambda t)$. When the model of §4.1 is generalized to include the concomitant effects of some other STD that is endemic, it is seen – after the dust has settled – that Λ is given by

$$1 = \frac{\beta}{\langle i \rangle} \left\langle \frac{i^2[(\Lambda+\mu+v+\gamma+ia\zeta)(\mu+\gamma)+(\Lambda+\mu+v+i\zeta+\gamma/a)i\zeta a^2}{(\mu+\gamma+i\zeta)(\Lambda+\mu+v)(\Lambda+\mu+v+\gamma+i\zeta)} \right\rangle. \tag{4.32}$$

Here β, a, Λ and v are as defined above, μ is the death rate from causes other than AIDS (note that in §4.1 we put $\mu = 0$), γ is the rate at which those infected with the other STD revert to an uninfectious but susceptible state (it may be that $\gamma = 0$), and ζ is the probability that a randomly chosen partner will infect a susceptible individual with this STD (analogous to the λ of §4.1 for HIV; ζ, in turn, may be calculated from the transmission parameters for the other STD, in the usual way). As in (4.8), the angled brackets denote averages over the initial distribution in rates of partner change, $\{p(i)\}$.

If the other STD has no effect on HIV transmission, $a = 1$ and (4.32) reduces to $\Lambda = \beta c - v - \mu$, with c defined by (4.14). This is simply (4.13) for Λ (except that $\mu = 0$ in §4.1). More generally, however, if the enhancing effects of other STDs are strong, $a \gg 1$, and if we assume $\gamma = 0$ (lifelong infection), (4.32) reduces to the rough approximation

$$\Lambda \approx \frac{\beta a^2}{\langle i \rangle} \left\langle \frac{i^3\zeta}{\mu+i\zeta} \right\rangle - v. \tag{4.33}$$

The factor a^2 here means that concomitant STD infection has a determining influence on HIV transmission. There are at least two strategies that can be employed here to reduce the rate,

Λ, at which HIV infection spreads. Programmes aimed at reducing rates of partner change, particularly among the most active groups, will as before reduce HIV transmission rates and will eventually reduce the prevalence of endemic STDs also, but the latter effect will be long- rather than short-term. Programmes aimed at treating other STDs can, if successful, increase the effective value of γ (the rate of converting to the uninfected state), and possibly also reduce the value of a (by reducing somewhat the incidence of symptomatic ulcers and chancres). Both these methods of reducing Λ by decreasing the incidence of other STDs will be more effective when a is large; neither will have any effect if $a = 1$.

5. Two-sex models for heterosexual transmission of HIV

Up to this point, all our models have been for HIV transmission in a single-sex population. For heterosexual transmission, two distinct populations must in general be considered, N_1 of males and N_2 of females. The male-to-female transmission probability, β_1, is thought to be higher than that for female-to-male, β_2, possibly by a factor of 2 or more (see figure 4 and table 2); it does appear that $\beta_1 > \beta_2$ for STDs such as gonorrhea, although this fact is of doubtful relevance to HIV (for a review of the data see May (1988)). The initial distribution in rates of partner-change, $\{p_1(i)\}$ and $\{p_2(i)\}$ for males and females respectively may also be different, subject to the obvious constraint that the mean rates must be equal, $m_1 = m_2$ (assuming the sex-ratio is initially 50:50 among the relevant age classes). As shown below, the average rate for male-to-female transmission is characterized by the parameter combination $\beta_1 c_1$, where c_1 is the mean-square to mean value for the males' partner-acquisition distribution (see (4.14)), and the corresponding female-to-male transmission rate is characterized by $\beta_2 c_2$. If these characteristic rates $\beta_1 c_1$ and $\beta_2 c_2$ are roughly equal, the system is symmetrical and we may collapse two-sex models back to effectively single-sex systems (note that males and females may have different distributions, so that c_1 and c_2 may be significantly different, even though the means must be equal). Thus there can be some justification for studying single-sex approximations to the heterosexual transmission of HIV.

In general, the overall transmission rate from males to females will not be identical with those from females to males. A two-sex version of the model of §4.1 is then

$$\mathrm{d}X_{\alpha,i}/\mathrm{d}t = B_{\alpha,i} - (\mu + i\lambda_\alpha)\, X_{\alpha,i}, \tag{5.1}$$

$$\mathrm{d}Y_{\alpha,i}/\mathrm{d}t = i\lambda_\alpha\, X_{\alpha,i} - (\mu + v)\, Y_{\alpha,i}. \tag{5.2}$$

Here the index α ($\alpha = 1$ for males, 2 for females) labels the two sexes, and demography along the lines of §3 has been introduced. $B_{\alpha,i}$ represents the rate at which new males or females enter the susceptible class with sexual activity level i. The probability that a male will acquire infection from any one female partner is

$$\lambda_1 = \frac{\beta_2 \sum i Y_{2,i}}{\sum i N_{2,i}}, \tag{5.3}$$

and vice versa for λ_2.

There are some new complications in the study of such two-sex models. As previously remarked, by ignoring any effects of age-structure and assuming the initial sex ratio is 50:50, male and female contacts will be in balance before the advent of HIV/AIDS so long as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

$m_1 = m_2$. But AIDS will not necessarily remove equal numbers of males and females, especially if transmission parameters differ for the two sexes. If this happens, the patterns in the distributions of acquiring new sexual partners must change over time, in such a way as to keep the total number of male and female contacts equal. Hyman & Stanley (1988) discussed this problem in more detail, and suggested a phenomenological modification of equations similar to (5.1) and (5.2) that preserves the male–female balance. A more complicated alternative is to work with the full distributions for males and for females acquiring new partners, and to allow these distributions to change over time in appropriate ways (which preserve the balance in contacts). Once it has been decided how to deal with this problem, the epidemiological dynamics can be explored numerically for any chosen set of parameters.

Yet again, interesting results can be gained by considering the early stages of the epidemic, when the equations may be approximately linearized and time-dependences characterized by $\exp(\Lambda t)$. Neglecting μ in comparison with v, (5.2) gives the expression $Y_{\alpha,i} \approx i\lambda_\alpha N_{\alpha,i}$. Substituting this into (5.3) gives an equation for λ_1 in terms of λ_2:

$$\lambda_1 = \beta_2 c_2 \lambda_2/(\Lambda + v). \tag{5.4}$$

Here c_2 is given by (4.14), with the averages taken over the distribution of rates of partner-change for females. A similar equation for λ_2 can be obtained by reversing the indices 1 and 2 in (5.4). The requirement that this pair of equations be consistent then gives the early growth rate of the epidemic as

$$\Lambda = (\beta_1 c_1 \beta_2 c_2)^{\frac{1}{2}} - v. \tag{5.5}$$

The corresponding ratio of incidence of HIV infection, and thus approximately of AIDS cases, among men to that among women can be seen to be $\lambda_1 m_1 N_1/\lambda_2 m_2 N_2$ in the early stages of the epidemic. But $m_1 N_1$ and $m_2 N_2$ must be equal (each sexual contact involves one man and one woman), so the early case ratio is λ_1/λ_2. Equations (5.4) and (5.5) then provide the approximate result

$$\frac{\text{HIV/AIDS among men}}{\text{HIV/AIDS among women}} \approx \left(\frac{\beta_2 c_2}{\beta_1 c_1}\right)^{\frac{1}{2}}. \tag{5.6}$$

It is often asserted that the roughly 1:1 ratio of AIDS cases among men and women in central Africa constitutes some kind of proof of heterosexual transmission. In fact, (5.6) shows there is no reason to assume such a 1:1 ratio if transmission efficiencies differ between men and women. Insofar as the male-to-female transmission probability, β_1, may be significantly larger than that for female-to-male, β_2, more cases might be expected among females in the early stages. But we explained in §2.5 that the variance in the partner-change distribution in Africa is thought to be significantly higher for females than for males, by virtue of the cadre of female prostitutes who preserve the overall balance in a population where males tend to be more promiscuous than the typical female. If this is so, then (see (4.15)) c_2 could be significantly larger than c_1. From an epidemiological standpoint, it is as if females on average had more sexual partners than males, because of the disproportionate role played by female prostitutes. Such effects could counter-balance any tendency for β_1 to exceed β_2, resulting in case ratios being roughly 1:1 for the two sexes. But any such 1:1 ratio is essentially a problem to be explained, not an automatic consequence of heterosexual transmission.

It is also possible to obtain expressions for the asymptotic fraction of each sex ever infected

as the epidemic spreads in a closed population, along the lines of §4.1.2. As could be deduced from (5.5), this analysis shows the basic reproductive rate for heterosexual transmission to be

$$R_0 = (\beta_1 c_1 \beta_2 c_2)^{\frac{1}{2}}/v. \tag{5.7}$$

That is, the parameter combination βc of the single-sex models is replaced by the geometric mean of $\beta_1 c_1$ and $\beta_2 c_2$ for the two separate populations. Further discussion of two-sex models is given by Hyman & Stanley (1988), Dietz (1988) and May *et al.* (1988*b*).

Not enough is known to make any reliable predictions about the average value of R_0 for heterosexual transmission in developed countries, but some very tentative remarks can be made. Suppose we take Peterman *et al*'s (1988) estimates to indicate that the transmission probabilities per partnership are very roughly $\beta_1 \approx 0.2$ and $\beta_2 \approx 0.1$; see figure 4 and table 2. Suppose also that distributions in rates of acquiring new sexual partners are similar for men and women in developed countries, so that $c_1 \approx c_2 \approx m + \sigma^2/m$ (where m is the mean, and σ^2 the variance, of the distribution in the numbers of new sexual partners acquired each year). Finally, suppose the duration of infectiousness, $1/v$, is roughly given by the average incubation time of 8–9 years. Thus a very rough approximation is $R_0 \approx (m + \sigma^2/m)$. That is, R_0 may exceed unity among heterosexual groups where new partners are typically acquired more often than annually, or for lower mean rates if the variance is large enough. This rough estimate is subject to the very important caveats given in §§2.3 and 4.3: if infectiousness varies significantly over the duration of the long and variable incubation interval, simple estimates based on expressions such as (5.7) can be misleading. If infectiousness is indeed effectively confined to two phases, each of about 1 year, at the onset of HIV infection and again at the onset of AIDS, then R_0 for heterosexual transmission is likely to require c-values exceeding one per year (equation (5.7) can, however, no longer be used).

6. Demographic consequences of HIV/AIDS

The basic model in §3 combined epidemiology with demography, to show that HIV/AIDS can reduce overall rates of population growth, and can even lead to population decline under some circumstances (see (3.12)). The basic model, however, was very simple. It ignored all age-structure (dealing with average birth and death rates per head), took no account of the possible effects of vertical transmission of HIV, and thence AIDS, to the offspring of infected mothers, and assumed that all HIV infections eventually produced AIDS. We now give a somewhat more detailed account of the possible demographic effects of HIV/AIDS in developing countries such as Africa, with models that incorporate age-structure and vertical transmission of HIV infection, and where a fraction f ($1 \geqslant f \geqslant 0$) of those infected go on to develop AIDS. The models, however, do retain the symmetry assumption that overall transmission rates male-to-female and female-to-male are roughly equal ($\beta_1 c_1 \approx \beta_2 c_2$), to provide an effectively single-sex population (Anderson *et al.* 1988; May *et al.* 1988*b*, *c*).

6.1. *A basic model with age-structure*

We define $N(a, t)$ to be the total number of individuals of age a at time t. These total numbers may, as before, be subdivided into susceptible, infected-and-infectious, and no-longer-infectious categories, $X(a, t)$, $Y(a, t)$, and $Z(a, t)$, respectively. The probability, per unit time,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

that a given susceptible will acquire infection is $c\lambda(a,t)$, which now depends explicitly on age; as in §3, we have neglected heterogeneity in degrees of sexual activity, and let c represent some appropriate average rate of acquiring new partners. We assume that death from causes other than AIDS occurs at the age-dependent rate per head $\mu(a)$, and births at the rate per head $m(a)$. All other rate processes are as in §3. The dynamical behaviour of this system is described by the following set of partial differential equations:

$$\partial X/\partial t+\partial X/\partial a = -[c\lambda(a,t)+\mu(a)]\,X(a,t), \tag{6.1}$$

$$\partial Y/\partial t+\partial Y/\partial a = c\lambda X-[v+\mu(a)]\,Y(a,t), \tag{6.2}$$

$$\partial N/\partial t+\partial N/\partial a = -\mu(a)\,N(a,t)-fvY(a,t). \tag{6.3}$$

These equations have as one boundary condition the requirement $X(0,t) = N(0,t) = B(t)$ and $Y(0,t) = 0$, where the birth rate, $B(t)$, is

$$B(t) = \int m(a)\,[N(a,t)-(1-\epsilon)\,Y(a,t)]\,\mathrm{d}a. \tag{6.4}$$

Here ϵ represents the probability that a child born to an infected mother will survive, whereas the remaining fraction $(1-\epsilon)$ die of AIDS in the first few years of life. As discussed in §2.1, ϵ is currently thought to be around 0.3–0.5 or more, although this number is not certain. The other boundary condition is given by specifying $X(a,0)$, $Y(a,0)$, and $N(a,0)$; that is, by specifying the age-specific numbers in each category at some initial time, $t = 0$.

To complete the description of this system, we generalize (3.5) to define $\lambda(a,t)$ as

$$\lambda(a,t) = \frac{\beta\int p(a,a')\,Y(a',t)\,\mathrm{d}a'}{\int p(a,a')\,N(a',t)\,\mathrm{d}a'}. \tag{6.5}$$

The ratio of integrals in (6.5) gives the probability that any one partner will be infected; $p(a,a')$ is the probability that a susceptible of age a will choose a partner of age a'. May *et al.* (1988*b*) have explored the asymptotic properties of this system of equations under the two extreme assumptions that all ages mix homogeneously ($p(a,a')$ is constant independent of a and a', so long as both ages lie in the sexually-active range) and that partners are restricted to the same age cohort ($p(a,a') = \delta(a-a')$). These two assumptions tend to represent opposite extremes, bracketing reality. May *et al.* found qualitatively similar results under these two extreme assumptions. In detail, the age-specific prevalence of HIV infection understandably tends to rise more slowly in the age-restricted model than in the homogeneously mixed one, resulting in HIV/AIDS having somewhat less demographic impact – other things being equal – if sexual pairings are age-restricted rather than homogeneously mixed. In what follows, we restrict attention to the homogeneously mixed case. For a more general analysis, see May *et al.* (1988*b*, *c*).

This system of age-structured, partial differential equations can now be solved numerically (Anderson *et al.* 1988). Starting with some specified set of age profiles for $N(a,0)$, $X(a,0)$, and $Y(a,0)$, and some specified set of demographic and epidemiological parameters, the initial birth rate, $B(0)$, and force of infection, $\lambda(a,0)$ are computed. Equations (6.1)–(6.3) then give the age profiles one time step later, and so on.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Figure 12 shows the results of one such computation, for a representative set of parameter values. In this example, the population continues to grow for some 50 years after AIDS first appears, but then begins slowly to decrease. As this happens, the age profiles change markedly, from the simple 'pyramidal' pre-AIDS profile to later more complicated profiles.

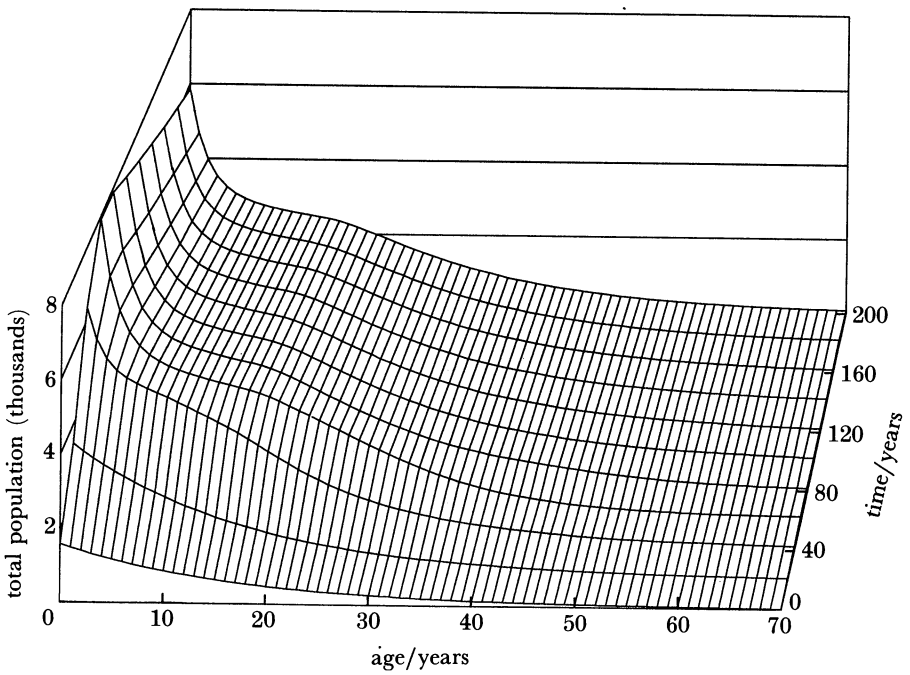

FIGURE 12. The changes that HIV/AIDS can produce in a population's age structure – numbers of individuals in the different age classes, from 0 to 70 years – are shown as a function of time (for the first 200 years after the onset of the epidemic). This illustrative example uses (6.1)–(6.5), with the assumption that partners are chosen at random within the sexually-active age-range (taken to be 15–50 years). The other demographic and epidemiological parameters in the model are: $\mu = 1/52$ per year; reproduction occurs at a constant rate for females between 15 and 50 years of age (the rate being such that $r = 0.04$ per year before AIDS); $v+\mu = 1/15$ per year; $\Lambda = \beta c - (v+\mu) = 0.233$ per year (see table 3); ϵ (survival probability for babies born to infected mothers) $= 0.3$; $f = 1$ (all HIV infections eventually lead to AIDS). The features of this figure are discussed in the text.

Figure 13 summarizes a collection of other examples, showing the dynamical behaviour of the population as a whole under various assumptions about f, the fraction of those infected who go on to die from AIDS. Again the long-term effects of HIV/AIDS can take several decades to show up (for a more detailed discussion, see May *et al.* (1988*b*, *c*)). Of course, on such long timescales many other things may happen. It could even be, as Lee (1987) and others have suggested, that 'homeostatic' mechanisms could come into play to resist such a decline.

6.2. *Asymptotic dynamics and age profiles*

As $t \to \infty$, the time dependence in $N(a, t)$ and other such quantities can be factored out, as $\exp(\rho t)$. As explained in May *et al.* (1988*b*), this is essentially because $\lambda(a, t)$ depends on the ratio of infecteds to total numbers, and thus becomes independent of time as $t \to \infty$. It follows that, after a sufficient length of time has elapsed, $N(a, t)$, $X(a, t)$, and $Y(a, t)$ will in general tend to exhibit stable age profiles, whose shapes do not change even though the total numbers increase or decrease.

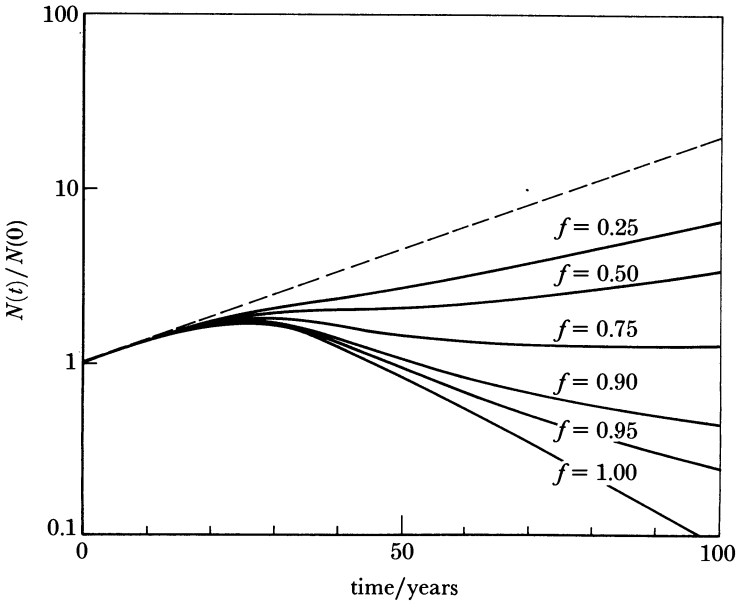

FIGURE 13. The magnitude of the total population is plotted (on a logarithmic scale, as a ratio to the initial magnitude) as a function of time, t, for a variety of f-values, as shown. The dashed line illustrated the AIDS-free rate of purely exponential growth; the curves have the features discussed in the text. These curves are derived from (6.1)–(6.5) with the epidemiological and demographic parameters having the values $\Lambda = 0.2$, $v = 0.06$, $\mu = 0.02$, $r = 0.03$ (all per year), $\epsilon = 0.5$; at $t = 0$, $Y(0)/N(0) = 0.01$.

We now analyse this asymptotic behaviour, and present numerical results for the asymptotically stable age profiles, and for social and economic indicators that can be derived from them.

The asymptotic age profiles can be obtained by first defining:

$$N(a,t) = n(a)\,G(a)\,e^{\rho t}, \tag{6.6}$$

$$X(a,t) = x(a)\,G(a)\,e^{\rho t}, \tag{6.7}$$

$$Y(a,t) = y(a)\,G(a)\,e^{\rho t}. \tag{6.8}$$

Here the function $G(a)$ is the stable age profile before the advent of AIDS, except that ρ replaces the pre-AIDS rate of population growth, r:

$$G(a) = \exp\left[-\rho a - \int_0^a \mu(s)\,ds\right]. \tag{6.9}$$

The functions $n(a)$, $x(a)$, and $y(a)$ describe the demographic effects of HIV/AIDS on the shapes of age profiles, as distinct from its effects on the overall growth rate ρ (and thus upon $G(a)$). Substituting (6.6)–(6.9) into (6.1)–(6.3), these functions obey the set of ordinary differential equations

$$dn/da = -fvy, \tag{6.10}$$

$$dx/da = -c\lambda x, \tag{6.11}$$

$$dy/da = c\lambda x - vy. \tag{6.12}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Equations (6.10)–(6.12) can now be integrated, to obtain analytic expressions for $n(a;\lambda,\rho)$, $x(a;\lambda,\rho)$ and $y(a;\lambda,\rho)$ as functions of age, a, and the parameters λ and ρ. The explicit asymptotic expressions for $N(a,t)$ and $Y(a,t)$, from (6.6) and (6.7), are substituted in (6.4) for $B(t)$ and (6.5) for $\lambda(t)$ to end up with two relations between the quantities λ and ρ. Eliminating λ, an explicit expression can in principle be obtained for the asymptotic rate of population growth, ρ, in terms of basic epidemiological and demographic parameters.

May *et al.* (1988*c*) have obtained such an explicit formula for ρ, under the simplifying assumptions that the death rate per head, μ, is constant and that the birth rate per head, $m(a)$, is the same for all ages above τ, where τ characterizes the onset of adult sexual activity:

$$\rho = -(v+\mu) + \nu\left(\frac{\beta c}{\beta c-\theta}\right)\left(\epsilon + \frac{[1-f]\,v}{\rho+\mu}\right). \tag{6.13}$$

Here ν is the overall average birth rate per head, which can be defined in terms of the pre-AIDS rate of population growth, r, as $\nu = (r+\mu)\exp[(r-\rho)\tau]$; May *et al.* (1988*c*) give an intuitive explanation of this result along with its derivation. The quantity $\theta \equiv fv + (1-\epsilon)\nu$ essentially represents the rate of adult and infant deaths from horizontally and vertically transmitted AIDS respectively. All the other parameters are as defined previously. Equation (6.13) embodies the further approximation that the upper age limit for sexual activity, ξ, is effectively infinite (some subtleties arise when the limits $f \to 1$ and $\xi \to \infty$ are both taken; these non-uniform limiting processes are discussed in May *et al.* (1988*c*)).

Equation (6.13) makes it plain that asymptotic rates of population growth can be driven negative by HIV/AIDS, depending on the relative magnitudes of the various epidemiological and demographic parameters. For a population whose growth rate per head is r before the advent of AIDS, the curves in figure 14 show the value of f – the fraction of those infected who go on to develop AIDS – that will eventually produce zero population growth ($\rho = 0$), for three

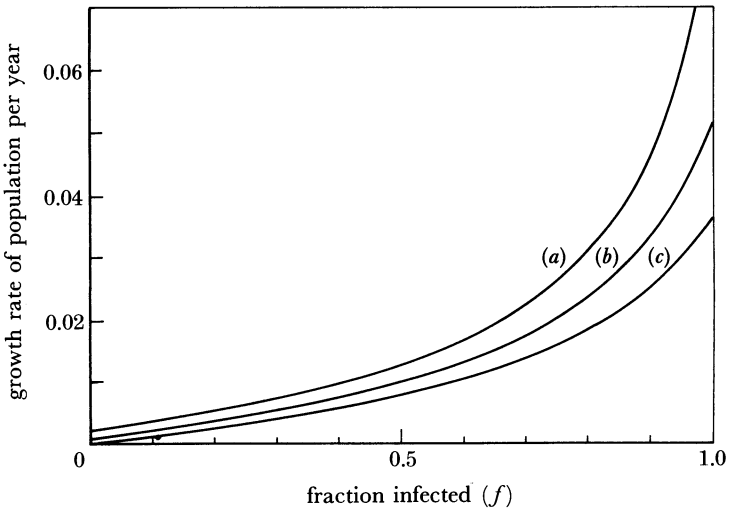

FIGURE 14. This figure shows the growth rate in the AIDS-free population, r, which can asymptotically be brought exactly to zero (stationary population) for a specified value of f. The other epidemiological and demographic parameter values are $\Lambda = 0.4$, $v = 0.1$, $\mu = 0.02$ (all per year), and, as in figures 12 and 13, within the age-range 15–50 years sexual partners are chosen randomly and reproduction takes place at a constant rate. The curves labelled (*a*), (*b*) and (*c*) are for $\epsilon = 0$, 0.5 and 1.0, respectively (ϵ is the survival probability for offspring born to infected mothers). These results are obtained from the critical version of (6.13) in which $\rho = 0$.

different assumptions about the survival probability for offspring born to infected mothers ($\epsilon = 0, 0.5, 1$). It appears that populations with reasonably large initial rates of population growth may eventually stop growing, if f is large enough.

Once ρ and λ have been found in terms of basic epidemiological and demographic parameters, the asymptotic age profiles for total numbers and numbers infected, $N(a) = n(a)\, G(a)$ and $Y(a) = y(a)\, G(a)$, respectively, can be calculated. Figure 15 shows such an asymptotic age profile in total numbers (and in numbers infected), and contrasts it with the pre-AIDS profile.

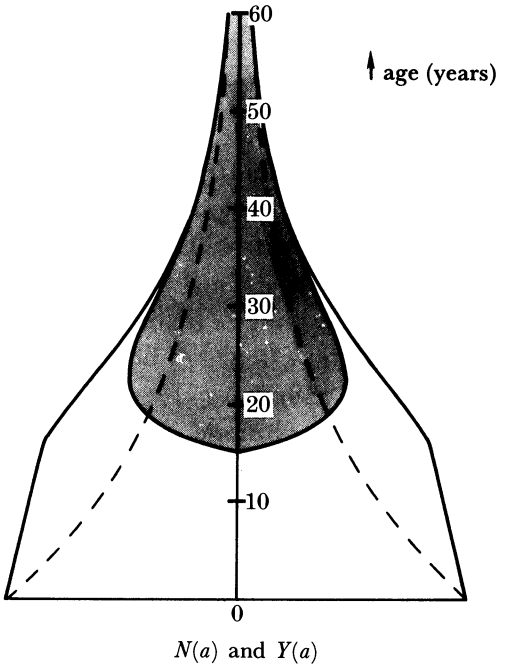

FIGURE 15. The asymptotic age profile of the total population, $N(a)$, and of the infected fraction, $Y(a)$ (the solid outer curves and the shaded inner region, respectively), are shown as functions of age, a (the vertical axis), under the assumption that sexual contacts are 'homogeneously mixed' among all age groups above 15 years. The demographic and epidemiological parameters here have the values $\Lambda = 0.2$, $v = 0.06$, $\mu = 0.02$, $r = 0.03$ (all per year), $\epsilon = 0.5$ and $f = 1.0$. The dashed age-profile is for the original, AIDS-free population, growing steadily at the rate $r = 0.03$ per year.

It is immediately apparent that deaths from horizontally and vertically transmitted HIV infections have two counteracting effects on the asymptotic form of the age profile for total numbers. On the one hand, increasing death rates and effectively decreasing birth rates cause the overall rate of population growth to decrease, which means – other things being equal – that age profiles tend to be less steep than is characteristically the case at present in developing countries. This effect, by itself, tends to decrease the ratio of numbers of children to numbers of adults. On the other hand, adult deaths from AIDS tend to steepen the age profile at older ages, and thus to increase the ratio of numbers of children to numbers of adults. It is not intuitively obvious which of these opposing tendencies – adult deaths tending to steepen the age profile or slowed population growth rates tending to make it less steep – will predominate for any specific set of demographic and epidemiological parameters.

One rough measure of, as it were, the ratio of tax consumers to tax producers in a developing

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

country is the 'child dependency ratio', CDR, which we define as the fraction of the total population below the age of 15 years. Figure 16 shows the value of this ratio, as a function of f, for several different values of the rate of population growth per head before AIDS (and a representative set of other demographic and epidemiological parameters). Here the opposing tendencies described in the previous paragraph roughly cancel, leaving the CDR essentially unchanged. That is, despite the pronounced changes in the asymptotic age profiles (see figures 12 and 15), the overall fraction below the age of 15 remains relatively unchanged.

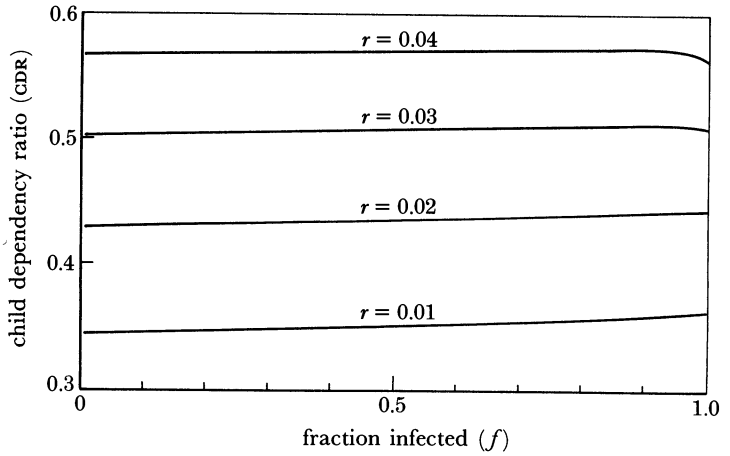

FIGURE 16. This figure shows asymptotic values of the 'child dependency ratio', CDR, defined here as the fraction of the total population who are below the age of 15 years. CDR is shown as a function of f (the fraction of HIV infectees who go on to develop AIDS), for several different values of the growth rate in the AIDS-free population, r, as shown. This figure is for $\Lambda = 0.4$, $\nu = 0.1$, $\mu = 0.02$ (all per year), and $\epsilon = 0.5$. As discussed in the text, CDR is roughly unaffected by HIV/AIDS for most values of the epidemiological parameters.

In short, the demographic effects of HIV/AIDS in developing countries are not simple. The analyses and examples presented above are based on grossly oversimplified models, but we believe they provide a basic understanding and a point of departure for more realistic numerical computations.

7. CONCLUSION

Although short-term predictions can sensibly be made by extrapolating current trends, long-term predictions about the prevalence of HIV and AIDS within any particular group requires an understanding of the nonlinear dynamics of transmission. The epidemiology of HIV has many unusual features. Among other things, more information is required about the distribution of incubation times for those who do develop AIDS, the fraction of those infected who will eventually go on to develop AIDS, the patterns of infectiousness over time, the overall transmission probability within a given kind of sexual or other relationship and the distribution in rates of acquiring new sexual or needle-sharing partners. The kinds of information that are currently available are surveyed in §2, partly for their own sake and partly as a basis for the epidemiological models developed in the rest of the paper.

Given the current uncertainties about so many biological and sociological aspects of HIV transmission, we believe it is sensible to begin by exploring relatively simple models that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

caricature the transmission dynamics, with a view to understanding qualitative features of the epidemic. Beginning with a very simple model, we have incorporated heterogeneities in levels of sexual activity, distributed incubation times and time-dependent patterns of infectiousness that are conditional on the incubation interval, changes in patterns of sexual behaviour over time, the possible effects of other STDs on HIV transmission and the interplay between HIV/AIDS and broad demographic processes.

The conclusions defy any brief summary. We have shown how the epidemiologically appropriate average number of new partners per unit time, c, can significantly exceed the simple mean of the distribution (§4.1), and how this possibly relates to the ratio of AIDS cases among men and women in Africa (§5). Rough estimates of the product of the transmission probability, β, multiplied by the average rate of partner-change, c, may be obtained from the doubling time of the epidemic in its early stages; these estimates are consistent with more directly derived estimates of β and c separately, although we emphasize that many such estimates may be compromised by significant time-dependence in patterns of infectiousness (§4.3).

The shape of observed epidemic curves for HIV seropositivity and AIDS incidence depart from those for more classical, measles-like, epidemics, in that early patterns of rapid growth soon give way to slower increases. We believe the observed patterns can be explained in terms of the substantial variabilities in rates of acquiring new partners within given risk-groups (§4.1). For populations having substantial heterogeneity in levels of sexual activity, the transmission dynamics of HIV will cause average rates of partner-change to fall over time, even if no individuals change their habits; this effect could interfere with some indirect methods for assessing changes in the behaviour of individuals with respect to rates of partner-change (§4.4).

Once fully established, HIV/AIDS is likely to have two counteracting effects on age profiles in developing countries. On the one hand, deaths from horizontally and vertically transmitted infections have the indirect effect of reducing rates of population growth, resulting in less steep age profiles (and a smaller fraction under the age of, say, 15). By contrast, the direct effect of adult deaths from AIDS makes for steeper age profiles among adults (and a higher fraction under the age of 15). Simple models suggest that these opposing tendencies may roughly cancel out, for representative values of the epidemiological and demographic parameters that may pertain in Africa; this conclusion is, however, very tentative (§6).

We have benefitted greatly from conversations with J. L. Aron, S. Blythe, C. Castillo-Chavez, J. M. Hyman, M. John, S. A. Levin, A. R. McLean, G. F. Medley, E. A. Stanley and many others. This work was supported in part by the Sloan Foundation and the NSF under grant DMS87-03503 (R. M. M.), and by the Overseas Development Agency, U.K., PANOS and the Medical Research Council (R. M. A.).

Appendix A

We sketch here the assumptions about the initial distribution, $\{p(i)\}$, and the consequent analysis, that underlie the illustrative figures 7, 8, 9 and 11.

As discussed more fully by Anderson *et al.* (1986), we assume that the initial distribution in rates of acquiring new sexual partners obeys a continuous gamma distribution, such that the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

proportion of the risk-group having between i and $i+di$ new partners per unit time is given by

$$p(i)\,di = \alpha^{\nu} i^{\nu-1} \mathrm{e}^{-i\alpha}\,di/\Gamma(\nu). \tag{A 1}$$

It follows that $\langle i\rangle = \nu/\alpha, \langle i^2\rangle = \nu(\nu+1)/\alpha^2$, and hence the variance of the distribution is $\sigma^2 = \nu/\alpha^2$. Rather than characterize this initial distribution by the parameters ν and α of (A 1), we prefer to use the parameters c and CV; here c is the effective average number of partners defined by (4.14) or (4.15), and CV represents the coefficient of variation of the distribution, $CV = \sigma/m$. These parameters c and CV are related to the α and ν of (A 1) by the relations

$$c = (\nu+1)/\alpha, \tag{A 2}$$

$$CV = \nu^{-\frac{1}{2}}. \tag{A 3}$$

The expectation values that enter into (4.9) and (4.10) can now be evaluated explicitly, to obtain

$$\langle 1-\mathrm{e}^{-i\phi}\rangle = 1-(1+\phi/\alpha)^{-\nu}, \tag{A 4}$$

$$\langle i(1-\mathrm{e}^{-i\phi})\rangle/\langle i\rangle = 1-(1+\phi/\alpha)^{-\nu-1}. \tag{A 5}$$

The dynamics of the epidemic are now described by (4.9), with the explicit form (A 5) for the relevant expectation value. The dependence of the dynamics on the basic epidemiological parameters can be seen more explicitly if $\phi(t)$ is replaced by $\psi(t)$, with the definition

$$\psi(t) \equiv (\nu+1)\,\phi/\alpha = c\phi(t). \tag{A 6}$$

The basic dynamical equation (4.9) then becomes

$$(\mathrm{d}\psi/\mathrm{d}t)\exp(-f\psi/R_0) = vR_0\{1-[1+\psi/(\nu+1)]^{-\nu-1}-[1-\exp(-f\psi/R_0)]/f\}+c\lambda(0). \tag{A 7}$$

The computational procedure is now straightforward. The epidemiological parameters R_0 (the basic reproductive rate), f (the fraction of HIV infectees who eventually proceed to develop AIDS), ν (related to the coefficient of variation of the initial sexual-activity distribution by $CV = \nu^{-\frac{1}{2}}$), and $c\lambda(0)$ (the magnitude of the initial infection that 'seeds' the epidemic) are first specified; the parameter v ($v = 1/D$, where D is the average incubation period) only enters in setting the timescale. Once these fundamental epidemiological parameters are specified, (A 7) can then be integrated to find $\psi(t)$ as a function of time, t.

All the other epidemiological quantities displayed in figures 7, 8, 9 and 11 can be calculated in terms of $\psi(t)$ and the parameters specified above. In particular, the fraction of the initial population to have experienced infection by time t, $I(t)$, is now given by

$$I(t) = 1-[1+\psi/(\nu+1)]^{-\nu}. \tag{A 8}$$

The cumulative number of AIDS cases, $C(t)$, is then obtained by integrating the first-order differential equation

$$\mathrm{d}C/\mathrm{d}t = v[fNI(t)-C(t)]. \tag{A 9}$$

Here N is the initial size of the risk-group in question, and $\mathrm{d}C/\mathrm{d}t$ is the rate at which new AIDS cases appear at time t. Seropositivity at time t, $S(t)$, is given by

$$S(t) = [NI(t)-C(t)]/[N-C(t)]. \tag{A 10}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Finally, the average number of new partners (per unit time) at time t, as given by (4.31), is

$$m(t) = \frac{m(0)\exp(-f\psi/R_0)}{1-C(t)/N} \tag{A 11}$$

Here $m(0)$ is the initial rate, $m = \langle i \rangle = \nu/\alpha$.

Figures 8, 9, and 11 are obtained by doing the above analysis. The asymptotic results in figure 7 are obtained more simply in the limit when $d/dt \to 0$, whereupon the relation between the asymptotic fraction ever to experience infection, $I(\infty)$, and the basic reproductive rate, R_0, can be seen to be given implicitly (for any specified value of f and $CV = \nu^{-\frac{1}{2}}$) by

$$I(\infty) = 1-[1+\psi_\infty/(\nu+1)]^{-\nu}, \tag{A 12}$$

$$R_0 = -f\psi_\infty/\ln(1-f\{1-[1+\psi_\infty/(\nu+1)]^{-\nu-1}\}). \tag{A 13}$$

References

Anderson, R. M. 1988 The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *Jl R. Statist. Soc.* A **151**, 66–98.

Anderson, R. M. & Johnson, A. M. 1988 Rates of sexual partner change in homosexual and heterosexual populations in the United Kingdom. In *Kinsey Institute conference on sexual behavior in relation to aids* (ed. B. Boller). New York: Oxford University Press. (In the press.)

Anderson, R. M. & May, R. M. 1979 Population biology of infectious diseases. *Nature, Lond.* **280**, 361–367; 455–461.

Anderson, R. M. & May, R. M. 1985 Vaccination and herd immunity to infectious diseases. *Nature, Lond.* **318**, 323–329.

Anderson, R. M., May, R. M. & McLean, A. R. 1988 Possible demographic consequences of AIDS in developing countries. *Nature, Lond.* **332**, 228–234.

Anderson, R. M., Medley, G. F., Blythe, S. P. & Johnson, A. M. 1987 Is it possible to predict the minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United Kingdom? *Lancet* i, 1073–1075.

Anderson, R. M., Medley, G. F., May, R. M. & Johnson, A. M. 1986 A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *I.M.A. J. Math. appl. Med. Biol.* **3**, 229–263.

BMRB 1987 *AIDS advertising campaign: report for surveys during the first year of advertising, 1986–87*. London: British Market Research Board.

Brookmeyer, R. 1988 A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J. Am. Stat. Ass.* (In the press.)

Burger, M. *et al.* 1986 Transmission of LAV/HTLVIII in sexual partners: seropositivity does not predict infectivity in all cases. *Am. J. Med.* **81**, 5–10.

Castillo-Chavez, C., Cooke, K. & Levin, S. A. 1988 The role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS). *Math. Biosci.* (In the press.)

Centres for Disease Control 1987 Antibody to human immunodeficiency virus in female prostitutes. *MMWR* **36**, 157–161.

Dietz, K. 1988 On the transmision dynamics of HIV. *Math. Biosci.* (In the press.)

Grant, R. M., Wiley, J. A. & Winkelstein, W. 1987 Infectivity of the Human Immunodeficiency Virus: estimates from a prospective study of homosexual men. *J. infect. Dis.* **156**, 189–193.

Hethcote, H. 1987 AIDS modeling in the USA. In *Future trends in AIDS*, pp. 35–41. London: HMSO.

Hethcote, H. W. & Yorke, J. A. 1984 *Gonorrhea transmission dynamics and control* (Lecture Notes in Biomathematics, no. 56). New York: Springer-Verlag.

HHS 1987 *HIV infections in the United States: a review of current knowledge and plans for expansion of HIV surveillance activities (a report of the Domestic Policy Council)*. Washington, D.C.: Department of Health and Human Services.

Hyman, J. M. & Stanley, E. A. 1988 A risk-based model for the spread of the AIDS virus. *Math. Biosci.* (In the press.)

Johnson, A. M. 1988 Social and behavioral aspects of the HIV epidemic: a review. *Jl R. Statist. Soc.* A **151**, 99–126.

Kermack, W. O. & McKendrick, A. G. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* A **115**, 700–721.

Kinsey, A. C. Pomeroy, W. B. & Martin, C. E. 1948 *Sexual behavior in the human male*. Philadelphia: Saunders.

Lee, R. D. 1987 Population dynamics of humans and other animals. *Demography* **24**, 443–465.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Macdonald, G. 1956 Theory of the eradication of malaria. *Bull. Wld Hlth Org.* **15**, 369–387.
May, R. M. 1987 Nonlinearities and complex behavior in simple ecological and epidemiological models. *Ann. N.J. Acad. Sci.* **504**, 1–15.
May, R. M. 1988 HIV infection in heterosexuals. *Nature, Lond.* **331**, 655–656.
May, R. M. & Anderson, R. M. 1987 Transmission dynamics of HIV infection. *Nature, Lond.* **326**, 137–142.
May, R. M., Anderson, R. M. & Johnson, A. M. 1988*a* The influence of temporal variation in the infectiousness of infected individuals on the transmission dynamics of HIV. In *AIDS 1988: AAAS Symposium papers* (ed. R. Kulstad). Washington, D.C.: American Association for the Advancement of Science. (In the press.)
May, R. M., Anderson, R. M. & McLean, A. R. 1988*b* Possible demographic consequences of HIV/AIDS: I, assuming HIV infection always leads to AIDS. *Math. Biosci.* (In the press.)
May, R. M., Anderson, R. M. & McLean, A. R. 1988*c* Possible demographic consequences of HIV/AIDS epidemics: II, assuming HIV infection does not necessarily lead to AIDS. In *Proceedings international symposium in mathematical approaches to ecological and environmental problem solving* (ed. C. Castillo-Chavez, S. A. Levin & C. Shoemaker). New York: Springer-Verlag. (In the press.)
May, R. M. & Jose, M. 1988 Epidemic propagation of sexually transmitted infections in behaviorally heterogeneous populations. *J. theor. Biol.* (In the press.)
Medley, G. F., Anderson, R. M., Cox, D. R. & Billard, L. 1987 Incubation period of AIDS in patients infected via blood transfusions. *Nature, Lond.* **328**, 719–721.
Medley, G. F., Billard, L., Cox, D. R. & Anderson, R. M. 1988 The distribution of the incubation period for the acquired immunodeficiency syndrome (AIDS). *Proc. R. Soc. Lond.* B **233**, 367–377.
Morse, P. H. & Feshbach, H. 1953 *Methods of theoretical physics.* New York: McGraw-Hill.
Pedersen, C., Nielsen, C. M., Vestergaard, B. F., Gerstoft, J., Krogsgaard, K. & Nielsen, J. O. 1987 Temporal relation of antigenaemia and loss of antibodies to core antigens to development of clinical disease in HIV infection. *Br. med. J.* **295**, 567–569.
Peterman, T. A., Stoneburner, R. L., Allen, J. R., Jaffe, H. W. & Curran, J. W. 1988 Risk of HIV transmission from heterosexual adults with transfusion-associated infections. *J. Am. med. Ass.* **259**, 55–63.
Quinn, T. C., Mann, J. M., Curran, J. W. & Piot, P. 1986 AIDS in Africa: an epidemiologic paradigm. *Science, Wash.* **234**, 955–963.
Reeves, G. & Overton, S. E. 1988 Average survival times for AIDS patients in the UK. *Lancet* i, 880.
Robertson, J. R., Bucknall, A. B. V. & Welsby, P. D. 1986 Epidemic of AIDS related virus (HTLVII/LAV) infection among intravenous drug abusers. *Br. med. J.* **292**, 527–529.
Ross, R. 1911 *The prevention of malaria* (2nd edn). London: John Murray.
Rothenberg, R. *et al.* 1987 Survival with Acquired Immunodeficiency Syndrome. *New Engl. J. Med.* **317**, 1297–1302.
Smith, C. E. G. 1970 Prospects for the control of infectious disease. *Proc. R. Soc. Med.* **63**, 1181–1190.
Wadsworth, J., Johnson, A. M., Elliott, P., Blower, S., Wallace, P., Miller, D., Adler, M. W. & Anderson, R. M. 1988 A pilot study of sexual lifestyles in a random sample of the U.K. population. *Br. med. J.* (In the press.)
Winkelstein, W., Lyman, D. M., Padian, N., Grant, R. M., Samuel, M., Wiley, J. A., Anderson, R. E., Lang, W., Riggs, J. & Levy, J. A. 1987 Sexual practices and risk of infection by HIV. *J. Am. med. Ass.* **257**, 321–325.

Discussion

M. E. IRWIN (*Department of Agricultural Entomology, University of Illinois, U.S.A.*). Professor May's models imply that HIV epidemics will progress along a 'natural' or 'biological' course. Yet his earlier comments gave credence to the emotional fear that accompanied, until rather recent times, plagues and other human epidemics. I believe that the human species is again feeling the pangs of fear. This leads me to hypothesize that the 'natural' or 'biological' pathway of HIV epidemics that he envisages may be drastically altered owing to people paying attention to educational advice on how to minimize the risk of contracting HIV. How might his model take these risk-evasive reactions into account?

R. M. ANDERSON. As quantitative information on changes in behaviour accrues, the model can be adapted to mirror such changes.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES