Transmission dynamics of HIV infection

Robert M. May and Roy M. Anderson

Simple mathematical models of the transmission dynamics of human immunodeficiency virus help to clarify some of the essential relations between epidemiological factors, such as distributed incubation periods and heterogeneity in sexual activity, and the overall pattern of the AIDS epidemic. They also help to identify what kinds of epidemiological data are needed to make predictions of future trends.

Fig. 1 The rise in seropositivity to HIV antigens in cohorts of patients over the period 1978–1985. The studies in San Francisco, London and New York were of homosexually/ bisexual males. The study in Italy is of drug addicts.

Exponential growth

When these assumptions are incorporated into a model for the transmission dynamics of HIV infection, the infected fraction of the population at risk (who are seropositive in tests for HIV) rises exponentially, as exp (λt), in the early stages of the epidemic. The exponential growth rate, λ, is related to the basic epidemiological quantities defined above by:

$$\lambda = \beta c - 1/D$$

The effective average over the distribution by degrees of sexual activity, c, is given explicitly as

$$c = \Sigma N_i/ \Sigma N = m + \alpha m$$

where m is the mean and α the variance of the distribution of the number of new sexual partners per unit of time. Thus, c is not simply the mean but the mean plus the ratio of variance to mean, which reflects the disproportionate role played by highly active individuals (in the tail of the probability distribution of sexual activity), who are both more likely to acquire infection and more likely to transmit it. The basic reproductive rate for HIV infection, R0, is related to the parameters β, c, and D, and hence to λ by the formula

$$R_0 = \beta c D$$

In contrast with standard epidemiological models in homogeneous populations (where the exponential phase of rising incidence lasts until something like half the pool of susceptibles have been infected), the early exponential phase is of
relatively short duration in our HIV models, giving way to a more nearly linear rise in the fraction infected (see Fig. 2).

This is because most susceptible individuals in the sexually highly active categories are infected in the early stages of the epidemic, producing saturation effects in these categories which decrease the exponential rise in incidence within them; although the incidence of infection continues to rise among individuals in less sexually active categories, the overall rate of increase is now slower than exponential.

Much less information is available about the rise in the number of individuals infected with HIV, as a function of time, than about the rise in the subsequent incidence of AIDS\(^{9,10}\), largely because information about infection requires serological examination for antibodies to the HIV virus. Although the initial infection may produce symptoms\(^{11,12}\) and, in some cases acute encephalopathy\(^{13}\) and meningitis\(^{14}\), it is not clear that such symptoms are always evoked: in any event, the symptoms are usually sufficiently mild to preclude systematic reporting. By contrast, the opportunistic infections characteristic of the destruction of the immune system in AIDS, leads to fairly reliable reporting\(^{15}\). There is, however, one study of hepatitis B virus (HBV) in a cohort of 6,875 homosexual and bisexual males in San Francisco, which resulted in serum samples being taken and preserved as early as 1978\(^{16}\), stored sera of a representative sample of 785 of these individuals gives the rise in the fraction seropositive for HIV, from 1978 to 1985,

<table>
<thead>
<tr>
<th>Area</th>
<th>Period</th>
<th>Serological data</th>
<th>Doubling time (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Male homosexuals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco, USA</td>
<td>1978–80</td>
<td>10–11</td>
<td></td>
</tr>
<tr>
<td>New York City, USA</td>
<td>1979–80</td>
<td>10–11</td>
<td></td>
</tr>
<tr>
<td>London, UK</td>
<td>1982–84</td>
<td>9–10</td>
<td></td>
</tr>
<tr>
<td>(b) Intravenous drug users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>1980–82</td>
<td>15–16</td>
<td></td>
</tr>
<tr>
<td>London, UK</td>
<td>1983–85</td>
<td>11–12</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1983–84</td>
<td>8–9</td>
<td></td>
</tr>
<tr>
<td>Case notifications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) All risk groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>1982–83</td>
<td>4–5</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>1983</td>
<td>15–16</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>1982–84</td>
<td>11–12</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>1981–83</td>
<td>9–10</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1982–84</td>
<td>13–14</td>
<td></td>
</tr>
<tr>
<td>Europe (EC)</td>
<td>1982–84</td>
<td>8–9</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>1982–84</td>
<td>8–9</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>1982–84</td>
<td>7–8</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>1982–83</td>
<td>7–8</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>1983–84</td>
<td>8–9</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>1983–85</td>
<td>9–10</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1983–85</td>
<td>9–10</td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1982–84</td>
<td>10–11</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>1982–84</td>
<td>9–10</td>
<td></td>
</tr>
<tr>
<td>West Germany</td>
<td>1982–84</td>
<td>6–7</td>
<td></td>
</tr>
<tr>
<td>(b) Heterosexuals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>1982–84</td>
<td>9–10</td>
<td></td>
</tr>
</tbody>
</table>

The pattern of roughly linear rise shown in Fig. 1 is characteristic of standard epidemics (in homogeneously mixed populations), but is suggested by our HIV models. In Britain and other countries in Europe, the virus seems first to have appeared several years later than in the United States (Fig. 2a, b and c), and the spread of infection is still in its early stages. As a result there are serological studies focused on HIV roughly from its initial appearance in Europe\(^{18,19}\). The initially exponential rise in HIV infection may be characterized by a doubling time, \(t_d\), related to the growth rate, \(\lambda\) of (1) by \(t_d = (\ln 2)/\lambda\).

Table 1 summarizes information about doubling times derived from serological and case notification studies, which lead to a surprisingly consistent estimate of \(t_d \sim 8–10\) months in the early stages of the epidemic (Fig. 2a) giving an estimate of \(A\) of about 1.0 yr\(^{-1}\). The characteristic duration of infection (and infectiousness), \(D\), is probably not significantly less than the characteristic time from HIV infection to manifestation of AIDS.

But \(D\) may be significantly longer if a substantial proportion of infected individuals remain asymptomatic carriers (with the epidemiology similar to hepatitis B virus\(^{20}\)). On the other hand, recent studies observing that measurable HIV antigen (HIV-Ag, the presence of which indicates the presence of the virus) appears early and transiently in primary HIV infection, that antibody production follows (1–3 months after infection) and that HIV-Ag may then disappear could imply lower estimates of \(D\), as would the apparent correlation of this persistence or reappearance of antigen with clinical, immunological and neurological deterioration\(^{1}\).

In the absence of conclusive data on infectiousness during the incubation period, we shall assume that \(D\) is equal to the incubation period. Studies of cases of AIDS associated with transfusion suggest that the average incubation period is 4–5 years\(^{1}\), but as such studies are extended, this estimate will rise (Fig. 3). The true average may be 8–10 years or more. Our estimate of \(A\) in conjunction with equation 1 then leads to the rough estimate

\[
\beta c = 1 \text{yr}^{-1} \quad (4)
\]

Note that \(A=\beta c\) provided \(D\) is large (4–5 years plus). Thus data on changes in seropositivity over time have allowed us to infer the approximate magnitude of the combination of epidemiological parameters \(\beta\) and \(c\), neither of which can easily be estimated directly.

Is this estimate consistent with what is known about \(\beta\) and \(c\) separately? Unfortunately, nearly all the information about degrees of sexual activity among male homosexuals has focused on average numbers of sexual partners, as distinct
the reported number of partners per unit of time may significantly overestimate the number of new partners per unit of time, which, or that equating $D$ to the incubation period, may overestimate the average duration of infectiousness. It may also be that the high values of $c$ arise from sampling biased towards the high activity groups of homosexual communities.

As public awareness about AIDS has increased, there have been changes in patterns of sexual activity among male homosexuals in the United States (reflected, for example, in marked decreases in the incidence of rectal gonorrhea\(^{24,25}\)) which have presumably resulted in changes both in $\beta$ and $c$. Our discussion, therefore, pertains mainly to the relatively early stages — 1978 to the early 1980s — of the epidemic.

**Incubation period**

Although much more information is available about the incidence of AIDS than of HIV infection (Table 1), it is harder to tease estimates of epidemiological parameters out of these. Incidence of AIDS depends not only on the transmission factors $\beta$ and $c$, but also on the incubation period and on the fraction, $f$, of those infected who will eventually develop AIDS.

Significantly, estimates of both the incubation period and $f$ have tended systematically to increase since the epidemic was first recognized\(^{5,26,27,28}\). Estimates of $f$ range from 10% to 75% or more\(^{29,30,31}\), with an incubation period of 4-5 years or more\(^1\). The progressive sequence of steps which eventually impair the ability of the immune system to respond to opportunistic infection seem not to be reversible. But whether all those infected with HIV are moving toward AIDS at different rates, or whether some will develop AIDS while others never will, remains unclear. Variability in the incubation period, and whether or not an infected person develops AIDS, could be accounted for by genetic heterogeneity within the host population (HLA-linked\(^{32}\)), or could be associated with specific strains of the antigenically variable HIV virus\(^3\).

Studies of the incubation period for those who develop AIDS suggest that the 'hazard function', the probability of the disease manifesting itself as a function of the time since infection, increases with time (Fig. 3). Lui and co-workers\(^4\) have assumed a Weibull distribution (a flexible two parameter probability distribution) for the incubation period with probability density function

$$h(t) = \gamma \exp[-(t/\alpha)^\beta]$$

(5)

If indeed the probability per unit time, to develop AIDS (for that fraction $f$ who do indeed develop it) increases linearly with time from infection as $\alpha t$, the result is a Weibull distribution with $\gamma = 2$ and $v = \alpha$ for the hazard function\(^*\). This

---

\* Although much more information is available about the incidence of AIDS than of HIV infection (Table 1), it is harder to tease estimates of epidemiological parameters out of these. Incidence of AIDS depends not only on the transmission factors $\beta$ and $c$, but also on the incubation period and on the fraction, $f$, of those infected who will eventually develop AIDS.

Significantly, estimates of both the incubation period and $f$ have tended systematically to increase since the epidemic was first recognized\(^{5,26,27,28}\). Estimates of $f$ range from 10% to 75% or more\(^{29,30,31}\), with an incubation period of 4-5 years or more\(^1\). The progressive sequence of steps which eventually impair the ability of the immune system to respond to opportunistic infection seem not to be reversible. But whether all those infected with HIV are moving toward AIDS at different rates, or whether some will develop AIDS while others never will, remains unclear. Variability in the incubation period, and whether or not an infected person develops AIDS, could be accounted for by genetic heterogeneity within the host population (HLA-linked\(^{32}\)), or could be associated with specific strains of the antigenically variable HIV virus\(^3\).

Studies of the incubation period for those who develop AIDS suggest that the 'hazard function', the probability of the disease manifesting itself as a function of the time since infection, increases with time (Fig. 3). Lui and co-workers\(^4\) have assumed a Weibull distribution (a flexible two parameter probability distribution) for the incubation period with probability density function

$$h(t) = \gamma \exp[-(t/\alpha)^\beta]$$

(5)

If indeed the probability per unit time, to develop AIDS (for that fraction $f$ who do indeed develop it) increases linearly with time from infection as $\alpha t$, the result is a Weibull distribution with $\gamma = 2$ and $v = \alpha$ for the hazard function\(^*\). This

---

\* Although much more information is available about the incidence of AIDS than of HIV infection (Table 1), it is harder to tease estimates of epidemiological parameters out of these. Incidence of AIDS depends not only on the transmission factors $\beta$ and $c$, but also on the incubation period and on the fraction, $f$, of those infected who will eventually develop AIDS.

Significantly, estimates of both the incubation period and $f$ have tended systematically to increase since the epidemic was first recognized\(^{5,26,27,28}\). Estimates of $f$ range from 10% to 75% or more\(^{29,30,31}\), with an incubation period of 4-5 years or more\(^1\). The progressive sequence of steps which eventually impair the ability of the immune system to respond to opportunistic infection seem not to be reversible. But whether all those infected with HIV are moving toward AIDS at different rates, or whether some will develop AIDS while others never will, remains unclear. Variability in the incubation period, and whether or not an infected person develops AIDS, could be accounted for by genetic heterogeneity within the host population (HLA-linked\(^{32}\)), or could be associated with specific strains of the antigenically variable HIV virus\(^3\).

Studies of the incubation period for those who develop AIDS suggest that the 'hazard function', the probability of the disease manifesting itself as a function of the time since infection, increases with time (Fig. 3). Lui and co-workers\(^4\) have assumed a Weibull distribution (a flexible two parameter probability distribution) for the incubation period with probability density function

$$h(t) = \gamma \exp[-(t/\alpha)^\beta]$$

(5)

If indeed the probability per unit time, to develop AIDS (for that fraction $f$ who do indeed develop it) increases linearly with time from infection as $\alpha t$, the result is a Weibull distribution with $\gamma = 2$ and $v = \alpha$ for the hazard function\(^*\). This
assumption differs from conventional epidemiological models, where infected individuals move through the incubation interval either at a fixed rate, or in a fixed time. But none of this resolves the question of what proportion of those infected will develop AIDS on what timescale. That issue will be resolved only by very long term (many decades) studies.

### Fraction eventually infected

In a closed and homogeneously mixed population, the total fraction eventually infected depends only on the basic reproductive rate of the infection, $R_e$, defined above as shown by the uppermost curve in Fig. 5. For sexually-transmitted infections such as HIV, the result can be extended to include the complications associated with a wide diversity in degrees of sexual activity.

In a closed population, the eventual fraction seropositive will depend both on $R_e$ and on the actual distribution of rates of acquisition of sexual partners. Assuming a gamma distribution, we may characterize it by $c$ and by its coefficient of variation ($CV = \sigma/m$). The resulting overall fraction infected is shown as a function of $R_e$, for a range of values of $CV$, in Fig. 5; for fixed $R_e$, the eventual seropositive fraction can be much lower than for $CV = 0$, if the variability in degrees of sexual activity (measured by $CV$) is high. This makes intuitive sense; the highly active individuals acquire infection, and eventually are removed, relatively early in the epidemic; transmission among the remaining, less active, individuals may be relatively weak.

Figure 5 may be used, in combination with two factual observations, to make a rough assessment of $R_e$ for HIV among male homosexuals. First, the studies indicate great variability in degrees of sexual activity among male homosexuals (with $CV$ significantly in excess of unity$^{12,22}$), thus confining attention to the lower curves. The second observation is that levels of seropositivity to HIV among male homosexuals in San Francisco in 1985 are variously reported as 70% or more$^1$ (in the HBV study which is probably biased towards more active individuals) and as around 50% (in a study carefully constructed to avoid bias$^5$), providing a lower bound of 50–70% on the proportion ever seropositive. For $CV$ noticeably in excess of unity, this can be achieved only if $R_e$ is in excess of 5.

Thus our early estimate of $R_e \sim 1$ yr$^{-1}$, in conjunction with the assumption that $R_e$ exceeds 5, leads to an indirect estimate that $D$ exceeds 5 years. Although $R_e$ like $\beta$ and $c$ changes with changing social and sexual habits, the data leading to our earlier estimate for $\beta$ come from the early stages of the rise in HIV infection, before such changes were significant. The estimate of $R_e$ depends importantly on observed levels of seropositivity, but these were also high before social changes became pronounced. Consequently, our estimate of $D$ which depends only on the recent biology of HIV, is reasonably consistent. This independent estimate of $D$ $\sim$ 5 years accords with current estimates that the incubation period is 4–5 years or more.

An estimate of the value of $R_e$ in the early stages of the epidemic is also valuable in indicating the magnitude of the social changes needed to bring $R_e$ below unity. If $R_e$ is around 5–10 or more, then reductions by a factor of 5–10 or more in $\beta$ are needed. Because $c$ depends disproportionately on those in the highly sexually active category, programmes aimed at getting them to change their habits — both to fewer partners and to "safe sex" — are most efficient. But if such individuals are less likely to respond to public health education, it will be harder to bring $R_e$ below unity.

### Mortality

The frequent assumption that the severity of the epidemic, in terms of cumulative mortality, will be greatest if all those infected eventually develop AIDS and subsequently die is not necessarily true. Mortality depends critically on the duration of infectiousness of both those infected who develop AIDS and those infected who do not. If the latter have a similar life expectancy to those not infected, but remain infectious for life, they may contribute more to the net transmission of the virus, $R_e$, than those who die of AIDS. Much may be understood by recognizing that the overall net reproductive rate of the virus, $R_e$, is made up of two components, the reproductive rate of those who develop AIDS ($R_{a}$) and the equivalent rate of those who do not ($R_{n}$). If a fraction $f$ develop AIDS

$$R_e = fR_a + (1-f)R_n$$

where the two reproductive rates are defined by equation (3) with different parameters for the separate groups. Even if the asymptomatic carriers are less infectious than those who develop AIDS, if they remain infectious over, say, a 30-year span of sexual activity, $R_e$ may be much larger than $R_{a}$, and, depending on $f$, the contribution of the asymptomatic carriers to $R_e$ may be dominant.

At present, it is not possible to tell whether the severity of the epidemic will be increased or decreased if a larger fraction of those infected develop AIDS, for the relative infectiousness of the two categories is unknown. For public health planning it is clearly important to attempt to acquire such data.

### Dynamics of the epidemic

The dynamics of an HIV epidemic within a homosexual community are represented by the results of our calculations given in Fig. 6, which shows the proportion seropositive and the incidence of cases of AIDS as a function of time since the start of the epidemic. It is assumed that 30% of those infected eventually manifest AIDS, with the incubation intervals obeying a Weibull distribution such that the average incubation period is 5 years$^3$. Individuals who are incubating AIDS are assumed infectious throughout the incubation interval, and the 70% who remain asymptomatic are assumed to remain infectious for similar periods.

Many of the features presented in Fig. 6 show qualitative agreement with observation. The rise in incidence of infection (seropositive) is initially exponential, but soon shows a more linear rise. And, the rise in incidence of AIDS lags that in the proportion infected, as seen.

It is easy to build epidemiological models of arbitrary complexity, which may appear beguilingly realistic, but we think there is little point in constructing them until more is known about the relevant epidemiological parameters. We distrust predictions made by using statistical procedures to fit polynomial or exponential curves to existing data on the incidence of AIDS, and then extrapolating$^{23}$. The HIV epidemic is a dynamic process; to predict future trends, models
Fig. 6 The predictions of a model (see ref. 8) incorporating variable incubation periods, heterogeneity in sexual activity and recruitment of susceptibles. The two graphs record changes in seroprevalence through time from the point of introduction of HIV into a community of 100,000 homosexual/bisexual males (graph a) and the incidence of AIDS yr$^{-1}$ (graph b). Heterogeneity in sexual activity is described by a gamma distribution with a mean fixed at 5 partners yr$^{-1}$ and variances 5, 25, 50 and 100 representing the predictions recorded by the four lines depicted in each graph. In a and b the smallest epidemic arises when the variance is largest and vice versa.

Parameter values, $R_0 = 5$, $D = 5$ yr, $f = 0.3$ with the life expectancy of AIDS patient set at 1 yr from diagnosis and for the susceptible sexually active community at 32 yr from the point of joining the sexually active class. The 70% of infecteds who do not develop AIDS are assumed to be infectious for a period equal to $D$. The immigration of new susceptibles into the sexually active community was set at 100,000 per 32 yr.

by purely heterosexual contact arises. The basic reproductive rate for such heterosexual transmission of HIV, $R_0^H$, is given by

$$R_0^H = (\beta_0 \delta) D$$

Here $\beta_0$ and $\beta$ are the transmission parameters for contacts between infected males and susceptible males and between infected males and susceptible females, respectively; $\delta$ and $\epsilon$ are as before given by (2), for the distribution in rates of acquiring new partners of the other sex by females and males, respectively.

Data are very limited on the transmission and sexual activity parameters, but the data in Fig. 4 suggest that $\delta$ and $\epsilon$ are significantly smaller than $c$ among homosexual males. Further, it seems likely that $\beta_0 < \beta$, and that both are less than the MM for homosexual males. Thus overall, the factor $(\beta_0 \delta) D$ seems likely to be much smaller than $c$ for homosexual males, which suggests that in developed countries, $R_0^H$ for purely heterosexual transmission is probably significantly smaller than $R_0$, for purely male homosexual transmission. Whether $R_0^H$ is greater than unity, such that HIV infection can maintain itself and spread by purely heterosexual transmission, is at present unclear. There is an urgent need for studies to measure $\delta$ and $\epsilon$, in different communities (stratified by age and social status) and to assess how these parameters change as a consequence of educational programmes and publicity campaigns on AIDS. The use of professional opinion poll organizations to gather quantitative data on rates of partner change over a series of specified time intervals by interview and questionnaire (Fig. 4d) could help to fill this gap in our knowledge, but estimates of $\beta_0$ and $\beta$ will come only from long-term studies of the heterosexual partners of infected patients.

If $R_0^H$ does exceed unity, the incidence of HIV infection in the heterosexual community will initially grow exponentially, at a rate given by the analogue of equation (1):

$$\Lambda = (\beta_0 \delta) D - 1/D = (R_0^H - 1) D$$

The estimates above indicate that initial doubling times will be significantly longer than the 9 months or so for HIV among homosexual males; the slow initial growth will be difficult to discern against a background of homosexual transmission among males and heterosexual transmission to females.

These observations are not necessarily inconsistent with the epidemiological situation for HIV in sub-Saharan Africa$^{14,15}$. In contrast to the United States and the United Kingdom, where male/female ratios of AIDS cases have been of the order of 14:1 to 20:1, in certain parts of central Africa including areas in Zaire, Rwanda and Uganda, sex ratios approaching unity have been reported$^{16,17}$. Very high prevalences of HIV antibodies have been found in males and...
females from surveys in urban and rural areas. These points suggest that heterosexual transmission has been frequent in both directions and horizontal studies have shown that infection is associated with the age-related degree of sexual activity amongst heterosexuals. We note, however, that in the early and approximately exponential phase of the epidemic, the ratio of the number of sero-positive males to seropositive females is not unity, but is roughly $\beta / c$. It is generally thought that $\beta$ is less than $\beta c$. Although the facts are uncertain (for gonorrhea, for instance, male-to-female transmission, $\beta$, is roughly twice $\beta$). Obviously the average number of heterosexual partners of females and males, $m$, and $m$, are equal, but $c$ could significantly exceed $c$, the variance of the distribution of rate of acquiring new sexual partners by females (associated with the concentrated activities of female prostitutes) is greater than that for males. This effect could partly offset $\beta$, being smaller than $\beta c$. Although there is no a priori reason to expect the ratio $\beta / \beta c$ to be exactly unity, its square root could easily be close to unity, which would explain the roughly equal proportions of seropositive males and females. Alternatively, the roughly equal proportions could be explained if homosexual transmission among males had coincidently raised the seropositive proportion among males to around the level among females, or by transmission by contaminated needles in public and private medical services. In any event, the rough equality of the seropositive proportions among males and females is a puzzle to be explained, and is not itself evidence for purely heterosexual sexual transmission.

Discussion

The ideas presented above are based on relatively simple mathematical models, with the aim of making clear some of the essential relations between epidemiological parameters and the overall course of HIV infection within various populations. Such models help to clarify what kinds of epidemiological data are needed to make predictions. As such data become available, the models can be made more detailed and realistic.

For public health planning, the dominant unknown is $f$, the fraction infected who will eventually develop AIDS. Estimates of this parameter have been increasing in recent years, but on present evidence the possibility cannot be ruled out that $f$ is less than 20% or as high as virtually 100%. Thus any current predictions about the number of homosexuals likely to acquire AIDS are uncertain by at least a factor 5 or so. Better understanding of the mechanisms of interactions between virus and host may help to determine $f$, but it is possible that only epidemiological data gathered on a decade-long timescale, as cases accumulate, will resolve this question.

The duration of infectiousness, and the way this duration is distributed among different infective, is also relevant to estimates of $R_{0}$ and hence the eventual number infected; more studies directed towards elucidating this information, including looking for virus in the blood, excretions and secretions of infected individuals over time, together with longitudinal studies of the patients of infected patients, are needed.

More generally, there is need for more studies that combine information about the epidemiological history of individuals with information about their sexual habits, such as the important study by Winkelstein et al. of an unbiased sample of homosexuals in San Francisco, which demonstrated the association between the number of sexual partners and probability of acquiring infection. We emphasize that what is epidemiologically important is the average rate of acquiring new sexual partners, not necessarily the same as the average number of partners per unit time. Some authors have recognized that sexually highly active individuals play a disproportionate role in the transmission dynamics, equation (2) quantifies this observation, making it clear that the epidemiologically relevant quantity is not the mean number of new partners but, rather, the mean-square divided by the mean.

In developed countries, at present and into the near future, it is probable that sexually-transmitted HIV infections among females are likely to come mainly from bisexual males. Whether subsequent spread of infection from such females to heterosexual male partners is likely to reach significant levels, and more importantly whether purely heterosexual transmission of HIV infection may be self-sustaining ($R_{0} > 1$), depends on estimates of the transmission parameters $\beta$, $c$, and $c$.

We have shown how $c$ for transmission among homosexual males can be estimated indirectly from data on initial doubling times, but corresponding estimates of $\beta c$ and $\beta$ are much harder, partly because the corresponding doubling rates are likely to be longer and partly because these infections are likely to be masked by homosexual/bisexual transmission among males, and by bisexual-to-female transmission among females (both of which processes depend simply on $c$). Attempts to estimate these quantities directly, and thence to estimate $R_{0}$ more accurately, are urgently needed.

From present knowledge, it is not possible to assess whether $R_{0}$ is greater or less than unity in developed countries, and it is thus not possible to say whether HIV infections could spread epidemiologically by purely heterosexual transmission. The evidence from Africa, however, clearly argues that the sexually active population as a whole should be regarded as at risk.

We have greatly benefited from discussions with Anne Johnson, Mike Adler, John Pickering, Graham Medley, Stephen Blythe and Jenny Crombie. Financial support from the MRC and the NSF is gratefully acknowledged. We thank C.A. Carne, I.V. Weller and T. McManus for permission to quote unpublished data (Fig. 4).

Robert M. May is in the Biology Department, Princeton University, Princeton, New Jersey 08544, USA. Roy M. Anderson is in the Parasite Epidemiology Research Group, Department of Pure and Applied Biology, Imperial College, University of London, London SW7 2BZ, UK.

---