INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460892-1

SMR.301/29

FIRST AUTUMN WORKSHOP ON MATHEMATICAL ECOLOGY

(31 October - 18 November 1988)

A PRELIMINARY STUDY OF THE TRANSMISSION DYNAMICS OF THE HUMAN IMMUNODEFICIENCY VIRUS (HIV), THE CAUSATIVE AGENT OF AIDS

R.M. ANDERSON
Imperial College
London, UK

IMA Journal of Mathematics Applied in Medicine & Biology (1986) **3**, 229–263

A Preliminary Study of the Transmission Dynamics of the Human Immunodeficiency Virus (HIV), the Causative Agent of AIDS

R. M. ANDERSON, G. F. MEDLEY

Parasite Epidemiology Research Group, Department of Pure and Applied Biology, Imperial College, University of London, London SW7 2BB

R. M. MAY

Biology Department, Princeton University, Princeton, N.J. 08544 U.S.A.

AND

A. M. JOHNSON

Academic Department of Genito-Urinary Medicine, The Middlesex Hospital Medical School, James Pringle House, The Middlesex Hospital, London W1N 8AA

[Received 23 October 1986 and in revised form 12 November 1986]

The paper describes some preliminary attempts to formulate simple mathematical models of the transmission dynamics of HIV infection in homosexual communities. In conjunction with a survey of the available epidemiological data on HIV infection and the incidence of AIDS, the models are used to assess how various processes influence the course of the initial epidemic following the introduction of the virus. Models of the early stages of viral spread provide crude methods for estimating the basic reproductive rate of the virus, given a knowledge of the incubation period of the disease (AIDS) and the initial doubling time of the epidemic. More complex models are formulated to assess the influence of variation in the incubation period and heterogeneity in sexual activity. The latter factor is shown to have a major effect on the predicted pattern of the epidemic; high levels of heterogeneity decrease its magnitude. Areas of biological uncertainty, future research needs, and public health implications are discussed.

Keywords HIV; AIDS; transmission dynamics; variable incubation periods; heterogeneity in transmission; basic reproductive rate of infection; doubling time of an epidemic; serological survey.

1. Introduction

BASIC RESEARCH on the biology of the virus responsible for acquired immunodeficiency syndrome in humans (AIDS) and the pathology associated with infection has proceeded at a rapid rate since the discovery of the aetiological agent in 1983 by Barre-Sinoussi *et al.* (1983) and Gallo *et al.* (1983). Competing

This paper was presented at the IMA Conference on "The Mathematical Theory of Biological Systems" held in Oxford, 7–9 July, 1986.

claims by the French and American research groups over who first isolated the causative agent of AIDS created some confusion over the correct terminology for the virus (named Human T-Lymphotrophic Virus type III (HTLV-III) by Gallo *et al.* (1983) and Lymphadenopathy Associated Virus (LAV) by Barre-Sinoussi *et al.* (1983). A recent report by the International Committee on Taxonomy of Viruses, however, endorsed the name 'human immunodeficiency virus' (HIV) to resolve the controversy regarding priority of discovery (Coffin, 1986). This new name describes the host and a major biological property of the virus.

A striking illustration of the pace of research in this area is provided by a recent paper by Hahn *et al.* (1986) that describes genetic variation in the AIDS virus (HIV) via genomic analysis, molecular cloning, and nucleotide sequencing. These authors compared the nucleotide, and deduced amino acid sequences of the gene encoding the extracellular envelope glycoprotein of the virus (the gene is called *env*) in four to six isolates obtained from each of three patients over a 1-year to 2-year period. Hahn *et al.*'s (1986) study highlights the hypervariability of *env* relative to the remainder of the viral genome, and pinpoints localized regions of hypervariability within the gene. Genetic changes among different viruses result largely from duplications, insertions, or deletions of short stretches of nucleotides, as well as from an accumulation of nucleotide point mutations. Such variability appears to be correlated with antigenicity and might also give rise to viruses with altered virulence or tissue tropism. The detail of such research must be viewed in the context of the relatively short period that has elapsed since the virus was first isolated (since 1983).

In marked contrast to our current understanding of the molecular biology and genetic structure of the aetiological agent is the present state of the epidemiological research that is concerned with virus transmission and persistence within human communities (Peterman, Dotman, & Curran, 1985; Anderson & May, 1986). Little is known at present concerning the basic epidemiological parameters that characterize virus transmission dynamics. Among the unknowns are such factors as the duration of the latent period of infection (the time from infection to the point when the host is infectious to other members of the population), the proportion of infected people who will go on to exhibit end-stage disease (i.e. 'full blown' AIDS), the duration of viral persistence within the host following infection, and the infectious period. The lack of knowledge in this area creates many difficulties in the design of effective control policies, in assessing future trends in the incidence of AIDS, and in planning provision of health-care facilities.

This paper describes a preliminary study of the transmission dynamics of HIV based on analyses of the quantitative epidemiological data that are available, and on the formulation of simple deterministic mathematical models of viral spread and persistence. It is organized as follows. Section 2 provides a brief background to past studies of the dynamics of sexually transmitted diseases (STDs) and the factors that distinguish the transmission dynamics of such infections from the more frequently studied directly transmitted viral infections such as measles. Section 3 examines the available epidemiological data and attempts to derive parameter estimates for the central processes that control viral transmission.

Sections 4 and 5 detail the formulation of various deterministic models of viral spread starting with the early stages of an epidemic (Section 4) and moving on to more complex formulation describing heterogeneity in exposure to infection and variable incubation periods (Section 5). Section 6 provides a general discussion of future research needs and current problems in model formulation and analysis.

2. Sexually transmitted diseases (STDs)

Most STDs have certain characteristics that cause their epidemiology to be somewhat different from infections such as measles and rubella (German measles), which have received the bulk of attention in the literature concerned with mathematical epidemiology (e.g. Anderson & May, 1985). First, for infections such as gonorrhoea, syphilis, and genital herpes, only sexually active individuals need to be considered as candidates in the transmission process. In contrast with simple 'mass action' transmission models for measles, a doubling of the hosts' population density does not tend to increase the rate at which new infections are produced by infectious people. Second, the carrier phenomenon, in which certain individuals harbour asymptomatic infection, is important for many STDs. In the case of gonorrhoea, for example, many women are virtually asymptomatic and so do not seek treatment and remain active spreaders of infection for relatively long periods of time. The carrier state may well be of particular importance in the transmission of the AIDS virus, since many individuals appear to be seropositive (they possess antibodies specific to the HIV antigens) but they do not show overt symptoms of infection.

Third, many STDs induce little or no acquired immunity following recovery from infection, so that most individuals who have been treated or who have spontaneously recovered rejoin the susceptible class (e.g. gonorrhoea). In the case of HIV, the situation is probably more complex since, even in those individuals who are seropositive but without symptoms of AIDS, the virus probably persists within the host. Viral persistence following HIV infections may well be lifelong, since the retroviruses (the group to which HIV belongs) are characterized by their ability for long-term persistence within the body of the host (Weiss, 1982). HIV has a complex life cycle that includes a chromosomally integrated proviral DNA stage that has the potential for indefinite persistence.

Fourth, the transmission of most STDs is characterized by a high degree of heterogeneity generated by great variability in sexual habits among individuals within a given community.

This set of characteristics—virtual absence of a threshold density of hosts for disease-agent persistence, long-lived carriers of infection, absence of lasting immunity, and great heterogeneity in transmission—give rise to infectious diseases that are well adapted to persist in small low-density aggregations of people.

Past work on mathematical models of STDs is largely centred on two infections, namely Hepatitis B and gonorrhoea. The work of Cooke & Yorke (1973) on models for gonorrhoea provided the impetus for a series of papers in

the late 1970s and early 1980s describing models which take account of the characteristics described above. The recent monograph of Hethcote & Yorke (1984) on the dynamics and control of gonorrhoea provides an excellent summary of this body of research and is exemplary for the way the models are grounded on data and how the conclusions are aimed at public health workers in a way that emphasizes the ideas and not the mathematical details.

This body of research provides a useful starting point for the development of mathematical models for AIDS. There is, however, a series of features of HIV transmission that are unique amongst STDs and hence necessitate the development of new models. These will be described in Section 4. First we turn to the epidemiological data that are currently available to guide model formulation and to derive estimates of the central parameters of transmission.

3. Epidemiological data

The last few years have seen an enormous explosion in the volume of published papers concerned with the biology and the epidemiology of the AIDS virus. In this section we focus on the small fraction of this literature that contains quantitative data on the course of infection within individual patients and the spread of the infection within communities.

3.1 *Infection and Disease*

Infection with HIV results in a number of immunological abnormalities of widely varying severities in different patients. In patients with severely impaired lymphocyte function, serious disease may result from infection by opportunistic parasites, the development of cancers, and other serious conditions. Not all those infected, however, show severe immunodeficiency, and hence it is important, in the context of disease surveillance, to define tightly (as far as is possible) the symptoms associated with a case of AIDS. The case definition is as follows. 'A disease, at least moderately indicative of a defect in cell-mediated immunity, occurring in a person with no known cause for diminished resistance to that disease' (CDC, 1982). The definition requires specific evidence for the opportunistic infection or cancer, but it does not require direct evidence of immunodeficiency. The diseases considered indicative of underlying immunodeficiency are many and varied (Peterman, Drotman, & Curran, 1985) but the most frequent in cases of AIDS are *Pneumocystis carini,* Systemic Candidiasis, and Kaposi's sarcoma.

3.2 *Risk Groups in Human Communities*

Persons at increased risks of acquiring HIV infection include homosexual and bisexual men, intravenous (IV) drug abusers, persons transfused with contaminated blood or blood products, heterosexual contacts of persons with HIV infection, and children born to infected mothers. HIV is transmitted through sexual contact, perinatal exposure to infected blood or blood components, and

TABLE 1
AIDS: 'high risk' groups (in rank order of importance as judged by total case reports in the USA up to January 1986)

Patient group	Number of cases
(1) Homosexual/bisexual men, not IV drug users	10,600
(2) IV drug users	2,766
(3) Homosexual/bisexual men, IV drug users	1,310
(4) Transfusion recipients	261
(5) Heterosexual contact with HIV positives	182
(6) Children with HIV positive parent(s)	175
(7) Haemophilia patients	124

perinatal transmission from mother to neonate (CDC, 1985c). Virus has been isolated from blood, semen, saliva, tears, breast milk, and urine, and is likely to be isolated from other body fluids, secretions, and excretions. Epidemiological evidence has as yet only implicated blood and semen in transmission. Studies of nonsexual household contacts of AIDS patients indicate that casual contact with saliva and tears does not appear to result in transmission. Transmission to normal household contacts of infected persons has not been detected when the household contacts have not been sexual partners or have not been infants of infected mothers. The kind of non-sexual person-to-person contact that generally occurs among workers, clients, or others in the workplace does not seem to pose a risk for transmission of HIV. The epidemiology of HIV infection is thus somewhat similar to that of Hepatitis B (HBV) infection. The high-risk groups for HIV infection are very similar to those for HBV infection (CDC, 1985c). An indication of relative risk between different groups of people is reflected in Table 1 which records the numbers of reported cases of AIDS in the USA up to 13 January 1986 in different groups such as homosexuals and IV drug abusers (CDC, 1986). Among the group at greatest risk (homosexual men), recent research suggests that frequent and receptive anal intercourse with many partners predisposes to HIV infection (Goedert *et al.*, 1984).

3.3 *Incidence of AIDS*

Since October 1980, when the first cases of AIDS were reported in the United States (some cases in the USA have since been traced back to 1978) the infection has rapidly spread to Europe and to many other regions of the developed world. The situation in Africa is more complex, and it is still uncertain how long the disease has been present on this continent. However, evidence is accumulating that heterosexual transmission is the predominant mode of spread. The incidence in many of these regions (defined as reported cases per annum) has risen exponentially with, as yet, no signs of turnover in the epidemic. The case reports, particularly in the USA, provide a clear picture of the magnitude and seriousness

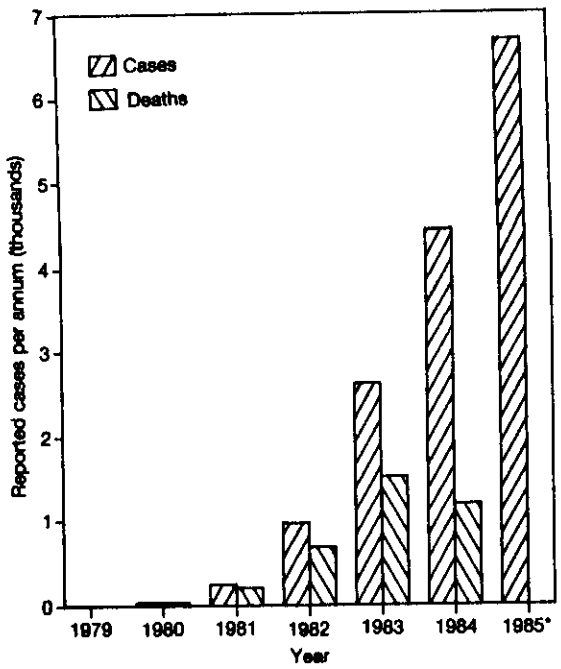

FIG. 1. Reported cases of AIDS per annum (incidence) and reported deaths from AIDS in the United States over the period 1979–1985 (the 1985 figure only includes cases up to November while the deaths for 1985 are unavailable at present). The reported deaths for the USA are thought to be an underestimate of the true figure (CDC, 1985a,b,c, 1986).

of the epidemic (Figs 1, 2, and 3). The doubling time in incidence, defined as t_d, differs among different countries and among different risk groups of people. However, as recorded in Tables 2 and 3, the variance around estimated doubling times for different countries is fairly small. On average, t_d is between 8 and 10 months in the early stages of the epidemic.

3.4 *Serology: Changes in the Proportion Infected with HIV Through Time*

Accurate serodiagnostic tests for HIV infection are now available. A series of cohort studies of different at risk groups (in particular, homosexual males, IV drug abusers, and transfusion patients) have provided good data on changes in the proportion seropositive through time. Data from two such studies, one in England (London) (Weber *et al.*, 1986; Carne *et al.*, 1985) and one in the USA (San Francisco) (CDC, 1985a) are recorded in Fig. 4. Doubling times for the proportion seropositive in the early stages of epidemics in different countries (Table 4) agree well with those derived from case reports (compare Table 4 with Table 2).

The relationship between the presence of antibodies specific to HIV antigens

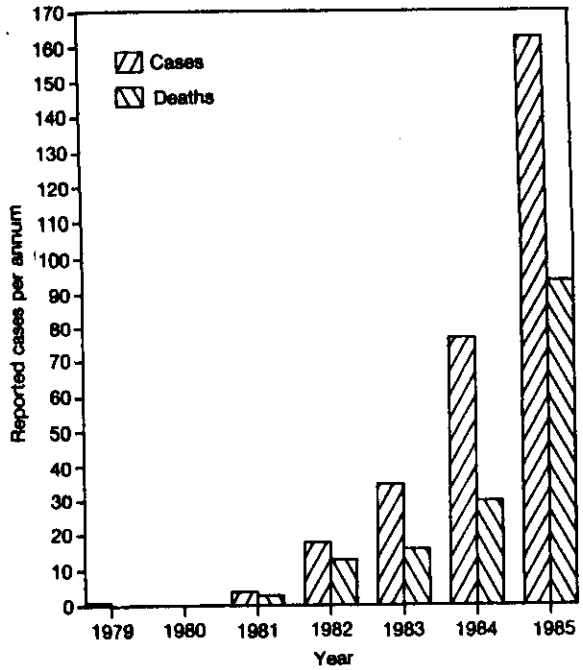

FIG. 2. Reported cases of AIDS per annum and deaths from AIDS in the United Kingdom over the period 1979–1985. Data from the Public Health Laboratory Service, Communicable Disease Surveillance Centre.

and the course of infection within an individual patient is unclear at present. Research on seroconversion in patients who have received infected blood products or transplant organs from infected patients suggest that antibodies are normally detectable between 40 and 60 days after infection (L'Age-Stehr *et al.*, 1985). Current thinking is that thereafter they will persist and be detectable for life. It will, of course, be many decades before this assumption can be rigorously tested by reference to long-term studies of infected patients. In the study of L'Age-Stehr *et al.* (1985) of four transplant patients who received kidneys from infected donors, antibody titres continued to rise for 12 months following the transplant of the infected organ.

Seropositivity seems to be almost invariably associated with the presence of the virus. Wong-Staal & Gallo (1985) reported that greater than 80% of seropositive patients possess HIV in their peripheral blood.

3.5 *Incubation Period*

The incubation period is defined as the interval between the point of acquisition of the infection and the point of appearance of symptoms of

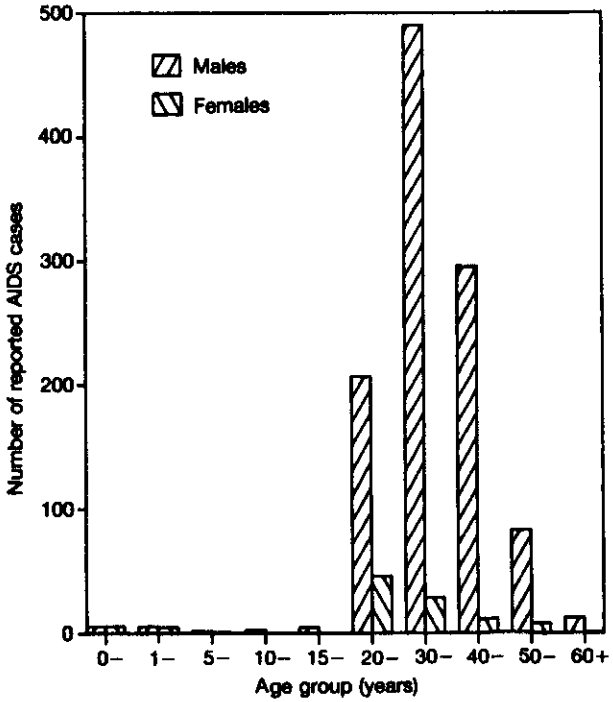

FIG. 3. The age and sex distribution of reported cases of AIDS in 18 European Countries up to June 1985 (Data source CDC, 1985b) (Austria, Belgium, Czechoslovakia, Denmark, Finland, France, Federal Republic of Germany, Greece, Iceland, Italy, Luxemburg, Netherlands, Norway, Poland, Spain, Sweden, Switzerland, United Kingdom).

TABLE 2
Doubling time t_d in AIDS incidence (in the early stages of the epidemic)

Country	Period	Doubling time (m)	*Rate* (r/yr)
Fed. Rep. Germany	1980–85	8·8	0·94
Australia	1983–85	4·8	1·73
Canada	1981–85	9·3	0·89
Austria	1983–85	15·6	0·53
Spain	1982–85	7·9	1·05
Sweden	1983–85	8·0	1·04
Switzerland	1983–85	9·9	0·84
Italy	1983–85	5·0	1·66
England	1982–85	6·6	1·27
USA	1981–85	9·2	0·90
Average		8·5	1·08

TABLE 3
Doubling time t_d in AIDS incidence by risk groups, USA (in the early stages of the epidemic)

Group	Doubling time (m)	Cases (1986)	% of total
Homosexual/bisexual men, IV drug abusers	9·5	599	6·9
Homosexual/bisexual men, non-IV drug abusers	9·1	5009	65·4
IV drug abusers	8·25	1429	16·5
Heterosexual contacts	7·85	100	1·1
Total	9·2	8661	100·1

Data from Peterman, Drotman & Curran (1985), CDC (1985a)

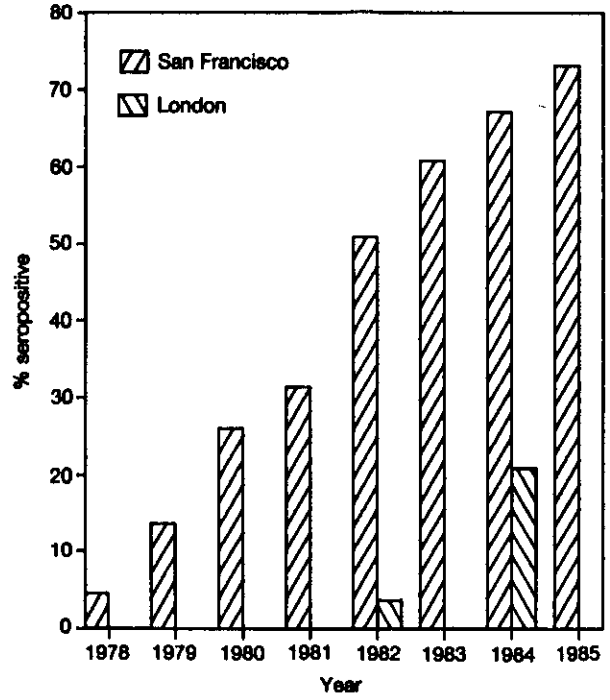

FIG. 4. The prevalence of antibodies to HIV antigens in a cohort of homosexual and bisexual men in San Francisco, U.S.A. over the period 1978–85 (CDC, 1985a). The figures for London in 1982 and 1984 are from a study of homosexuals attending a STD clinic (Carne *et al.*, 1985).

TABLE 4
Doubling time t_d from cohort HIV serology in the early stages of the epidemic

Country/City	Cohort	Period	Rate (r/yr)	Doubling time (m)
San Francisco	Homosexual	1978–80	0·73	11·3
New York	Homosexual	1982–83	0·77	10·7
London	Homosexual	1982–84	0·84	9·9
London	IV drug	1983–85	0·72	11·5
Switzerland	IV drug	1983–84	0·96	8·6
Italy	IV drug	1980–83	0·55	15·2

Data from CDC (1985a), Carne *et al.* (1985), Mortimer *et al.* (1985)

full-blown AIDS. Studies of blood-transfusion-associated cases of AIDS in the United States have provided good data on the distribution of the incubation period in fairly large samples of patients. The studies of Peterman, Drotman, & Curran (1985) and Lui *et al.* (1986), for example, focused on 194 cases of possible transfusion-associated AIDS. The distribution of the incubation periods in this sample is displayed in Fig. 5. The mean of this distribution is 30 months, but a high variance is shown in this sample of patients.

A simple deterministic model describing the incubation of AIDS is as follows. Let $y(t)$ denote the proportion of a cohort of patients (all of whom were infected with HIV at time $t = 0$) who have AIDS at time t. If we assume that the rate of conversion from seropositivity to 'full blown' AIDS is $v(t)$ at time t from the point of infection, then the rates of change of $y(t)$ and $x(t)$ (the proportion who do not have AIDS; $x + y = 1$) are given by

$$\frac{dx}{dt} = -v(t)x(t), \qquad \frac{dy}{dt} = v(t)x(t). \tag{3.1a,b}$$

The initial condition is, of course, $x(0) = 1$ and $y(0) = 0$. We assume for simplicity that all infected members of the cohort eventually develop AIDS. A simple assumption concerning the form of the function $v(t)$ is that it is linear with an intercept at zero ($v(t) = \alpha t$). A rationale to support this assumption would be that the progressive impairment of the patient's immune system through time from the point of infection with HIV results in a linear rise with time in the probability that an opportunistic infection or cancer develops in that patient. With these assumptions the solutions of equations (3.1) are

$$x(t) = \exp(-\tfrac{1}{2}\alpha t^2), \qquad y(t) = 1 - x(t). \tag{3.2a,b}$$

As illustrated in Fig. 5, the fit of this very simple model to the data of Peterman, Jaffe *et al.* (1985) (presented as incidence of AIDS in the sample of transfusion-infected patients per month) is remarkably good. Interestingly, equation (3.2a) is identical to the 'hazard function' of the Weibull distribution with probability

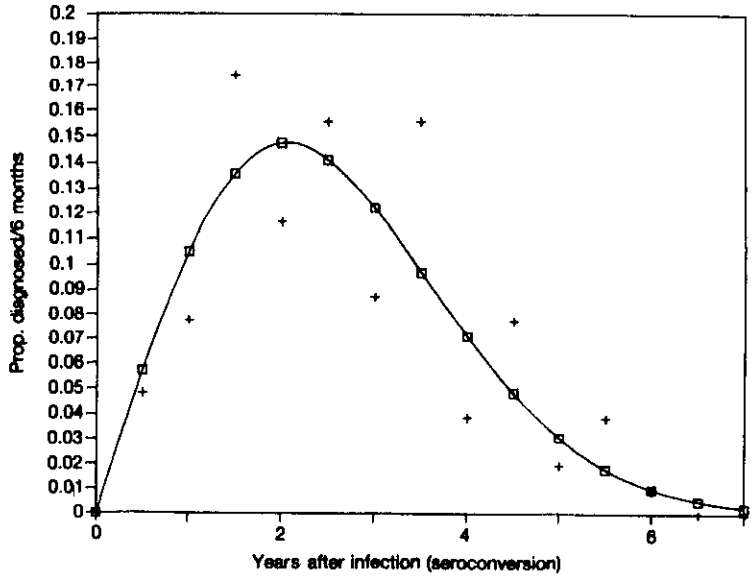

FIG. 5. The proportion of diagnosed cases of AIDS (proportion of all cases diagnosed over a 10-year period) in each 6-month interval from the point of seroconversion for antibodies to HIV (data from Peterman, Drotman, & Curran, 1985). The crosses denote observed values and the squares denote values predicted by the model defined in (3.2) in the main text with $\alpha = 0{\cdot}237\ \mathrm{yr}^{-1}$. Note that these data give a biased (over-) estimate of α (see Lui *et al.*, 1986).

density function

$$f(t) = a_1 a_2^{a_1} t^{a_1 - 1} \exp[-(ta_2)^{a_1}]$$

with the parameters a_1 and a_2 equal to 2 and $(\frac{1}{2}\alpha)^{\frac{1}{2}}$ respectively (Cox & Oakes, 1984). The Weibull distribution provides an excellent empirical description of the data recorded in Fig. 5, as has recently been noted by Lui *et al.* (1986). However, the simple model defined by (3.1) provides a more biologically orientated description of the observed pattern and may thus be of more general application in the formulation of models of the transmission dynamics of HIV.

3.6 *The Proportion of Infecteds who Develop AIDS*

A number of different studies of cohorts of infected patients from various high-risk groups have provided information about the proportion of infecteds that develop AIDS. At present, these studies only cover periods of 3 to 7 years and hence the proportions derived from them are probably underestimates of the true picture. As documented in Table 5, the estimates range from around 34% to as low as 8%. In the most extensive study, namely that of 6875 homosexual and bisexual men who attended a San Francisco City Clinic (USA), roughly one third of men infected over 5 years have developed AIDS (CDC, 1985a). This is an

TABLE 5
% Seropositive who develop AIDS over a three-year period (data from Goedert et al. (1986))

Study Group	Percentage
USA, Manhattan: Homosexuals	34·2
USA, Washington: Homosexuals	17·2
Denmark: Homosexuals	8·0
USA, Queens: IV Drug Abusers	12·5
USA, Hersley: Haemophilia Patients	12·8

indication that a fraction of those infected may not necessarily develop severe immunodeficiency as a result of viral invasion. The reasons for this pattern are unclear at present, although the recent study of Hahn *et al.* (1985) of genetic variation in HIV over time in patients with AIDS may suggest that different strains of the virus induce different severities of symptoms of disease. One particular aspect of this study is especially remarkable: genetic variation in viruses isolated at different times from the same patient is less than that observed between viral strains isolated from different patients (Hahn *et al.*, 1986). This seems puzzling in light of the observation that the 35 patients who formed the basis of this study were homosexual men from HIV endemic regions who had hundreds, and in some cases thousands, of different sexual partners over 1 to 2 year periods. On the one hand, extensive genomic heterogeneity of independent HIV viruses is common, yet, on the other hand, each patient appeared to be infected with only a very limited number of predominant viral forms. It is conceivable that immunological or non-immunological events that occur after the initial infection with HIV lead to protection from subsequent HIV infections (Hahn *et al.*, 1986).

3.7 *Survival of AIDS Patients*

Once AIDS is diagnosed, the survival period of patients is relatively short, on average being a matter of a few months to a few years. The mean survival period is commonly quoted as being between 9 and 12 months (Peterman, Drotman, & Curran, 1985) although this depends to some extent on the risk group and the opportunistic infection or cancer that is acquired by the patient. Patients with Kaposi's sarcoma have a better prognosis than patients with opportunistic infections (Moss *et al.*, 1984; Marasca & McEvoy, 1986). The median survival from diagnosis in the study of Marasca & McEvoy was 21 months for 44 cases presenting with Kaposi's sarcoma, and twelve months for 124 cases with other diseases.

3.8 *Latent and Infectious Periods*

Relatively little is known at present concerning the latent and infectious periods of HIV infection. The course of virus expression and antibody response from the time of infection through the incubation period and during different stages of clinical disease have not yet been clearly established (Esteban *et al.*, 1985). Patients with AIDS appear to have less virus in their blood than 'healthy' carriers or people with AIDS-related complex (Wong-Staal & Gallo, 1985). The serum level of antibody to HIV is high in most patients by the time clinical symptoms are recognized, and these antibody titres are sometimes reduced to low and barely detectable levels in advanced stages of disease (Salahuddin *et al.*, 1984). The period that elapses from initial infection to the point where an infected person is infectious to other susceptibles (the latent period) is thought to be of the order of a few days to a few weeks. Virus is detectable in blood, semen, and other secretions and excretions, a matter of days to weeks after initial infection (data from transfusion-associated cases).

Whether the presence of detectable levels of virus infection indicates infectiousness, however, is not clear at present. Similarly, the duration of infectiousness is not yet known. For those patients who proceed to develop full-blown AIDS (with an average incubation period of around 4–5 years) the infectious period is probably a little less than the mean incubation time (to account for a short latent period). Once symptoms of AIDS are diagnosed, such patients are effectively removed from circulation and hence cease to contribute to viral transmission within the community at large. In the case of infected (serologically positive) but apparently 'healthy' patients (certain of whom have been seropositive for periods approaching 8 years) the situation is more uncertain. As stated earlier, seropositivity implies persistent viral infection (probably lifelong) and hence such individuals do in principle constitute a pool of potentially infectious people. However, it may be that their infectiousness is somewhat less than individuals who go on to develop 'fully fledged' AIDS.

3.9 *Sexual Activity*

There are many distinct subgroups within the total population at high risk of HIV infection. Even within such subgroups, such as male homosexuals, the risk of acquiring infection will depend in part on the behaviour of an individual. In the following section, our attention is focused on the largest high-risk group, namely, homosexual males. In this sector of the population, the primary factors that determine the likelihood of acquiring infection appear to be the degree of sexual activity (defined as the number of different sexual partners per unit of time) and the nature of sexual practice (Goedert *et al.*, 1984). Data on these factors are understandably sparse at present and rarely quantitative in nature.

Two unpublished studies of sexual activity amongst homosexuals in London and one published study of a similar at risk group in San Francisco provide some data on the frequency distribution of the different sexual partners per unit of time (C. Carne & I. Weller, unpub; T. MacManus, unpub; McKusick *et al.*, 1985). This information is presented in Figs 6 and 7 and Table 6. A striking feature of the most detailed quantitative study (C. Carne & I. Weller unpub.) is the high average of the frequency distribution (defined as the mean number of different partners per month) and the very large variance (variance $\gg$ mean) (Table 6).

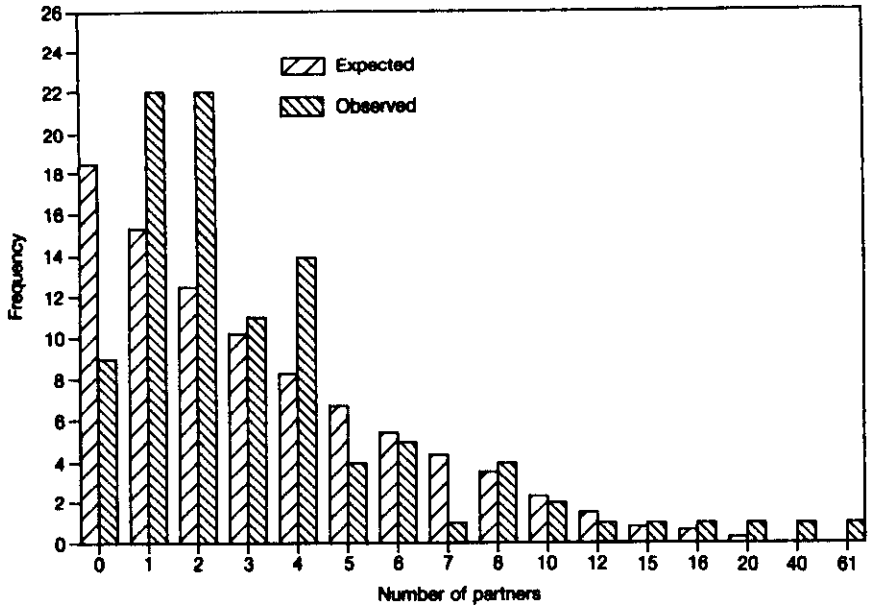

FIG. 6. Frequency distribution of the number of sexual partners per month among a sample of homosexual males attending a London STD clinic in January, 1986 (I. Weller & C. Carne, unpublished data obtained from interviews). The members of the sample were highly selected and came from a high-risk group (see main text). (Mean = 4·28 m^{-1}, variance = 57·9). The expected values are from the best-fit negative-binomial probability model which provides a poor fit to observed trends.

The data were collected, however, from a highly selected sample of patients with either persistent generalized lymphadenopathy (PGL) or AIDS, partners of patients with PGL or AIDS, or homosexual men with more than 10 partners in the last 3 months. As such, the mean is probably much higher than the mean of a randomly drawn sample from the total homosexual population of London. However, it is interesting to note that a similarly high mean was recorded by McKusick *et al.* (1985) in the San Francisco study (Table 6). This latter study also revealed evidence for a decrease in sexual activity over the period November 1982 to November 1986 as a result of wide publicity of the factors that determine the risk of acquiring HIV infection (Table 6).

The frequency distribution recorded by Carne & Weller is poorly described by the negative binomial probability model (Fig. 6). A slightly better, although not satisfactory fit is obtained with the gamma distribution (Fig. 8).

4. Models with homogeneous mixing

This section and the next two turn to the formulation of a series of mathematical models to mirror the transmission dynamics of HIV infection within

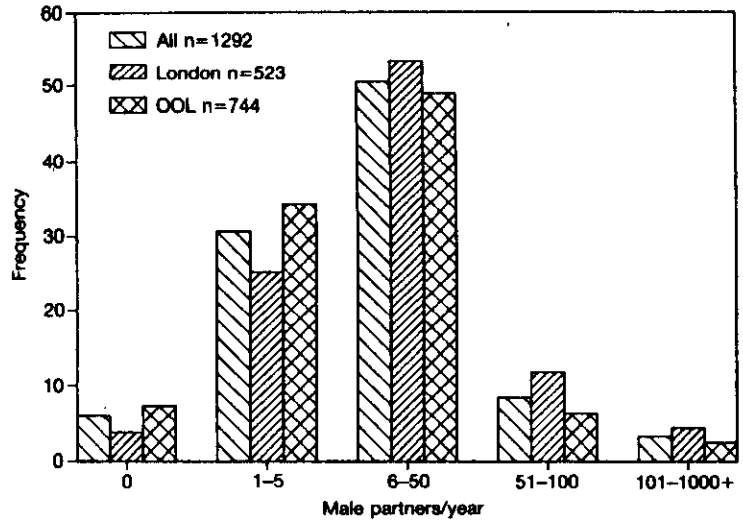

FIG. 7. Frequency distribution of the number of sexual partners of homosexuals (per year) drawn from London and outside London (OOL) (T. MacManus, unpublished data).

communities of homosexual males. By restricting attention to this single group we keep the models relatively simple, while still accounting for 70–80% of the known cases (Table 1).

We start by considering a very simple framework in which the population mixes homogeneously before moving on to more complex heterogeneous – mixing models in the following section. The reason for adopting this oversimplified framework initially is to help clarify how various observed factors influence the dynamics of disease spread and persistence. Throughout this and the next section we assume that: (1) susceptibles acquire infection via sexual contact with infectious people; (2) patients with AIDS are effectively withdrawn from circulation in the population such that they do not generate new cases of HIV

TABLE 6
Sex partners/month for homosexual males attending STD Clinics

	San Francisco, USA		London, England	
Month/Year	Mean	Variance	Mean	Variance
Nov 1982	6·8	—	—	—
May 1983	—	—	—	—
Nov 1983	4·8	—	—	—
May 1984	3·9	—	—	—
Nov 1984	2·6	—	—	—
1985	—	—	4·7	56·7

Data from McKusick *et al.* (1985), Weller & Carne (unpub.)

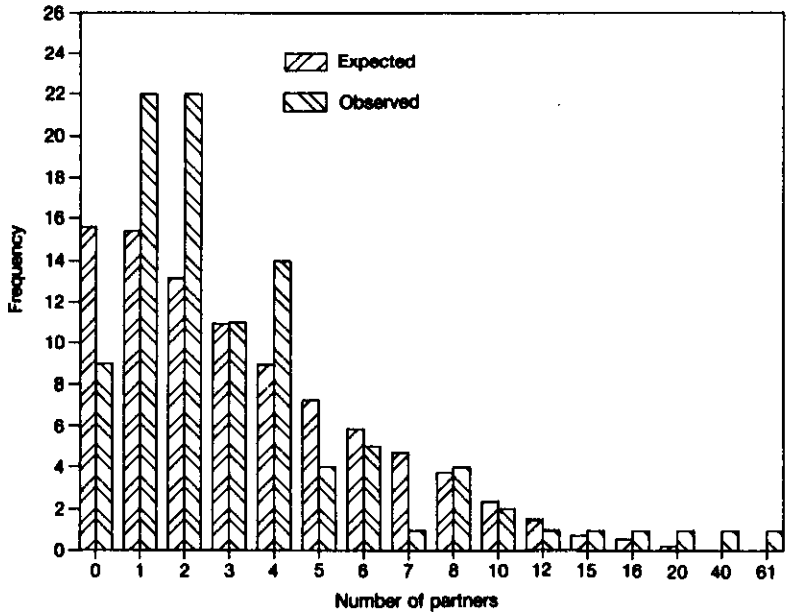

FIG. 8. The fit of the gamma distribution to the frequency distribution data of Weller & Carne presented in Fig. 6.

infection; (3) infected people are infectious for a period $1/v$ time units, after which a proportion p proceeds to develop AIDS while the remaining fraction $1-p$ passes into a 'seropositive' but non-infectious class, and (4) the latent period of infection is so short in relation to the incubation that it can be ignored. These assumptions are portrayed diagrammatically in a flow chart in Fig. 9a.

4.1 *The Early Stages of the Epidemic*

We first consider a closed population of fixed size N in which the density of susceptibles and infectious people (homosexual men) at time t are denoted by $X(t)$ and $Y(t)$ respectively. We ignore most of what happens following infection (i.e. whether individuals progress to AIDS or not) and simply denote the rate of movement out of the infectious class as v, where $1/v$ denotes that average incubation period, such that a proportion p of those leaving the infectious class (e.g. pvX) enters the AIDS patient class. AIDS-related deaths are ignored in the early stages of the epidemic.

The model is defined as

$$\frac{dX}{dt} = -\lambda cX, \qquad \frac{dY}{dt} = \lambda cX - vY. \tag{4.1a,b}$$

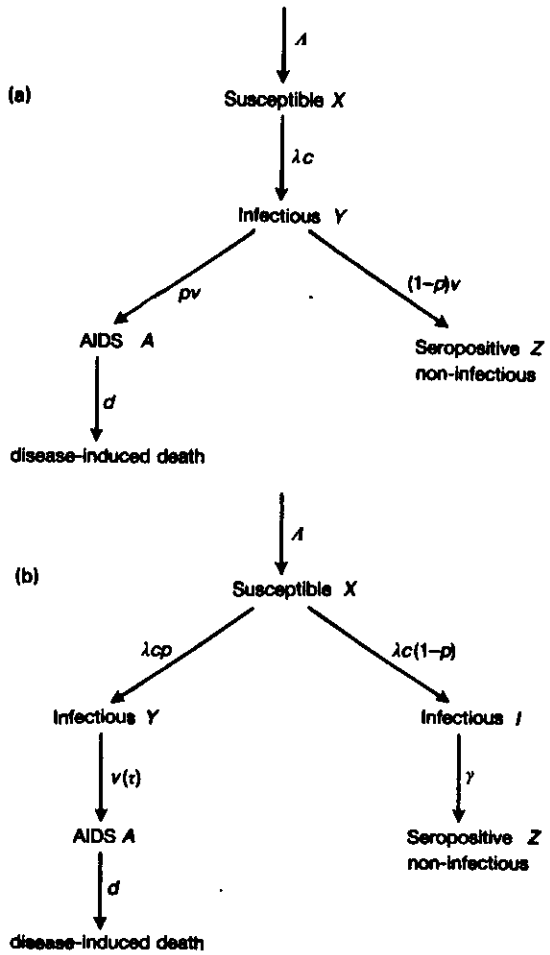

FIG. 9. (a) Flow diagram of the different model structure described in (4.1) in the text. Assumes homogenous mixing and a constant rate, v, of leaving the incubating (infectious) class. (b) Equations (4.11) divide the incubating/infectious class into two subgroups (those who develop AIDS and those who do not) in order to introduce a variable incubation period, $v(\tau)$. Note that all individuals in each subgroup suffer a background mortality rate μ.

Here, c is the average number of sexual partners (the nature of this average is investigated in more detail in the following section) and λ is the probability of acquiring infection from a randomly chosen partner. The term is proportional to the transmission probability β and to the probability Y/N of a given partner being infected:

$$\lambda = \beta Y/N. \tag{4.2}$$

In the early stages of the epidemic, $X \approx N$ and (4.1b) and (4.2) give

$$\frac{dY}{dt} \approx (\beta c - v)Y. \tag{4.3}$$

That is, the number of infectious people is given by

$$Y(t) \approx Y(0) \exp[(\beta c - v)t], \tag{4.4}$$

where the incidence of diagnosed AIDS is simply $pvY(t)$.

Equation (4.3) gives the doubling time t_d of the epidemic during its early stages:

$$t_d = (\ln 2)/(\beta c - v). \tag{4.5}$$

For this very simple model, the basic reproductive rate R_0 of HIV infection is defined as the number of secondary infections produced, on average, by one primary infection in a wholly susceptible population. From (4.3), R_0 is given as

$$R_0 = \beta c D, \tag{4.6}$$

where D is the average incubation period; $D = 1/v$. Equation (4.5) may therefore be expressed in terms of R_0, where

$$t_d = (\ln 2)D/(R_0 - 1) \tag{4.7}$$

Clearly, $R_0 > 1$ for the infection to trigger an epidemic. Equation (4.6) gives an explicit formula for the change in average sexual habits among homosexuals that is needed to reduce transmission of HIV infection below this threshold: the distribution of sexual habits needs to change such that the new mean activity $\hat{c}$ obeys

$$\hat{c} < c/R_0. \tag{4.8}$$

The main function of this very simple model is to provide a crude method of estimating the magnitude of R_0 given information on the doubling time t_d (in seropositives or cases of AIDS—see Tables 2 and 4) of the epidemic during its initial phase of growth. Equation (4.7) suggests that R_0 lies in the range 5–6 for a doubling time of around 9 months and an incubation period in the range 4–5 years. More generally, (4.5) gives $\beta c \simeq 1\ \text{yr}^{-1}$, independent of the estimate of v (provided $D \approx$ several years); then we have the rough relation $R_0 \approx D$ (with D measured in years).

4.2 *The Epidemic in a Population with Recruitment*

The model defined in Section 4.1 above can be extended to consider the full course of the epidemic in a population of size $N(t)$ at time t subject to immigration at a rate Λ and a natural mortality rate μ (in the absence of infection the steady state population size $N^* = \Lambda/\mu$). We assume that those with AIDS die at a rate d (where $1/d \approx 9$ months $-$ 1 year). We denote susceptibles, infectious, AIDS patients, and non-infectious seropositives as $X(t)$, $Y(t)$, $A(t)$, and $Z(t)$, at time t, respectively.

$$\frac{dX}{dt} = \Lambda - \mu X - \lambda c X, \qquad \frac{dY}{dt} = \lambda c X - (v + \mu)Y, \tag{4.9a,b}$$

$$\frac{dA}{dt} = pvY - (d + \mu)A, \qquad \frac{dZ}{dt} = (1 - p)vY - \mu Z \tag{4.9c,d}$$

where λ is as defined in (4.2) and $N(t) = X(t) + Y(t) + A(t) + Z(t)$. Analytical and numerical studies reveal that, provided $R_0 > 1$, the system settles to a steady state with HIV infection maintained within the population. The approach to this equilibrium, however, is oscillatory in character, where the interepidemic period T is given very approximately (see Anderson & May, 1982) by

$$T \approx 2\pi[(\hat{L}/R_0)D]^{1/2}. \tag{4.10}$$

Here, D is the average incubation (= infectious) period and $\hat{L}$ is the sexually active life expectancy of a homosexual in the absence of HIV infection once he has joined the homosexually active population (say 18–50 years). With R_0 values in the range of 4–5 and with D set at 4·5 years, equation (4.10) (or, more generally, with the rough relation $R_0 \approx D$ noted at the end of the previous section) gives the very crude prediction of damped epidemics every 30–40 years. But the parameters characterizing social behaviour (such as A and c, and thence R_0) are unlikely to remain unchanging on such a time scale.

Numerical studies of the model defined by (4.9) give a clearer picture of the pattern of the initial epidemic following the introduction of HIV into a totally susceptible homosexual community. A summary of the parameter values used in these studies is presented in Table 7; the population size was set at 100 000 with a life expectancy ($1/\mu$) of 32 years in the absence of infection. Two simulations of the trajectories of AIDS incidence (/annum/10 000) and the proportion seropositive ($y(t) = [A(t) + Y(t) + Z(t)]/N(t)$) in the population through time are recorded in Figs. 10 and 11. Note that the model predicts that incidence will reach its maximum value 12–15 years after the arrival of the infection (with R_0 values in the range 4–5.0). Also note that the predicted change in the proportion

TABLE 7
HIV infection, crude parameter values

Parameter	Estimate
Latent period	Days to a few weeks
Incubation period	4·5–5·0 years (+?)
Infectious period	4·5–5·0 years (?)
Seroconversion	A few weeks
Life expectancy of AIDS patient	9 months – 1 year
Doubling time (initial)	8 – 10 months
Proportion seropositive who develop AIDS	10% – 30% (+?)
Average sex partners/month	2 – 6
Variance in sexual habits	Variance/mean ratio, 5–20?

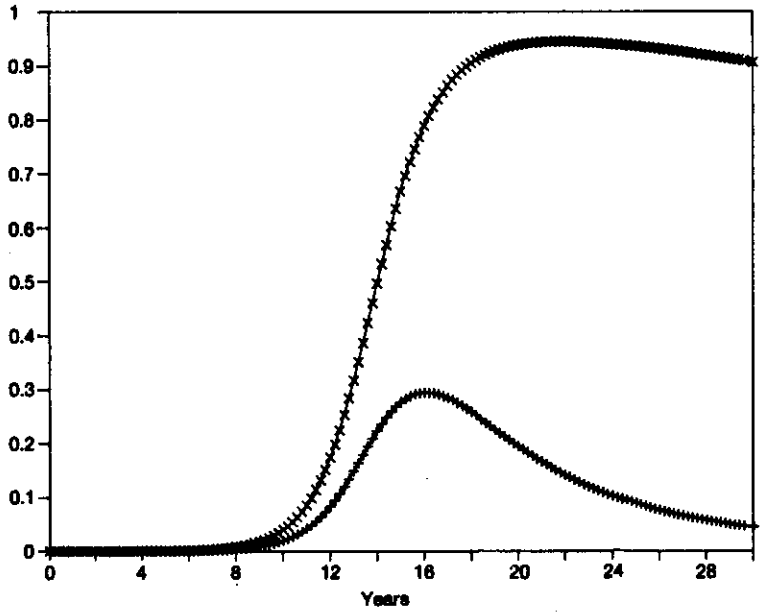

FIG. 10. Temporal predictions of the homogeneous-mixing model with recruitment of susceptibles ((4.9) in the text). The graph denotes temporal changes in the proportion seropositive (×) and the incidence of AIDS (+) (incidence/year/10 000). Parameter values: $R_0 = 5{\cdot}15$, $D = 5$ yr, $d = 1\,\text{yr}^{-1}$, $p = 0{\cdot}3$, $N(0) = 100\,000$, $\mu = (1/32)\,\text{yr}^{-1}$, $\Lambda = 1333{\cdot}3\,\text{yr}^{-1}$.

seropositive over the first 8 years is similar to that observed (compare Fig. 4 with Figs. 10 and 11). In short, despite the simplicity of the model and its many shortcomings, predicted patterns crudely mirror observed trends in homosexual communities.

4.3 *Variable Incubation Periods*

As discussed in an early section that examined the available epidemiological data, the incubation period of AIDS varies widely between infected individuals. It was shown (see Fig. 5) that a reasonable approximation of the observed pattern is given by a model in which the incubation period is depicted as a linear function of the time τ that a person has been incubating the infection ($v(\tau) = \alpha\tau$, see (3.1*a*)). The model defined in (4.9) can be easily modified to take account of this biological property. We define $Y(t, \tau)$ as the number of infectious people who will develop AIDS that have been incubating the virus for τ time units at time t, and $I(t)$ as the number of infectious people who will not develop AIDS (they are assumed to have an average duration of infectiousness of $1/\gamma$ time units) (see Fig.

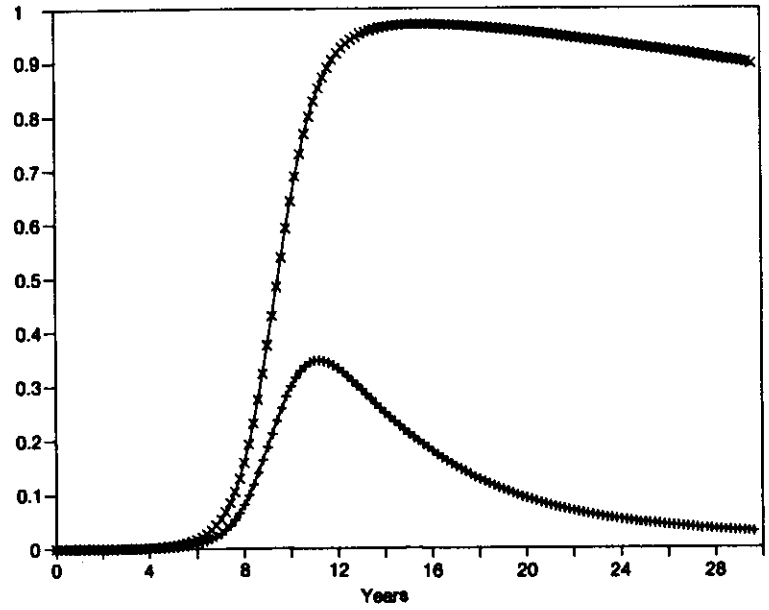

FIG. 11. Identical to Fig. 10 except $R_0 = 7{\cdot}0$.

9b). The new model is

$$\frac{dX}{dt} = \Lambda - \mu X - \lambda c X, \qquad \frac{\partial Y}{\partial t} + \frac{\partial Y}{\partial \tau} = -v(\tau)Y - \mu Y, \tag{4.11a,b}$$

$$\frac{dI}{dt} = (1-p)\lambda c X - (\mu + \gamma)I, \tag{4.11c}$$

$$\frac{dA}{dt} = \int_0^t v(\tau)Y(t, \tau)\,d\tau - (d + \mu)A, \qquad \frac{dZ}{dt} = \gamma I - \mu Z. \tag{4.11d,e}$$

Here, $\lambda = \beta\bar{Y}/N$, where

$$\bar{Y}(t) = \int_0^t Y(t, \tau)\,d\tau + I(t),$$

and the initial and boundary conditions of (4.11b) are defined as

$$Y(t, 0) = p\lambda c X, \qquad Y(0, \tau) = \begin{cases} \hat{Y} & \text{for } \tau = 0, \\ 0 & \text{for } \tau \neq 0, \end{cases}$$

with $\hat{Y}$ the initial number of infectious people (=1 throughout the paper). A clear picture of the influence of a variable incubation period is provided via numerical studies of the model defined in (4.11), and the comparison of these predictions with those of the simpler model in which the incubation period v is assumed to be

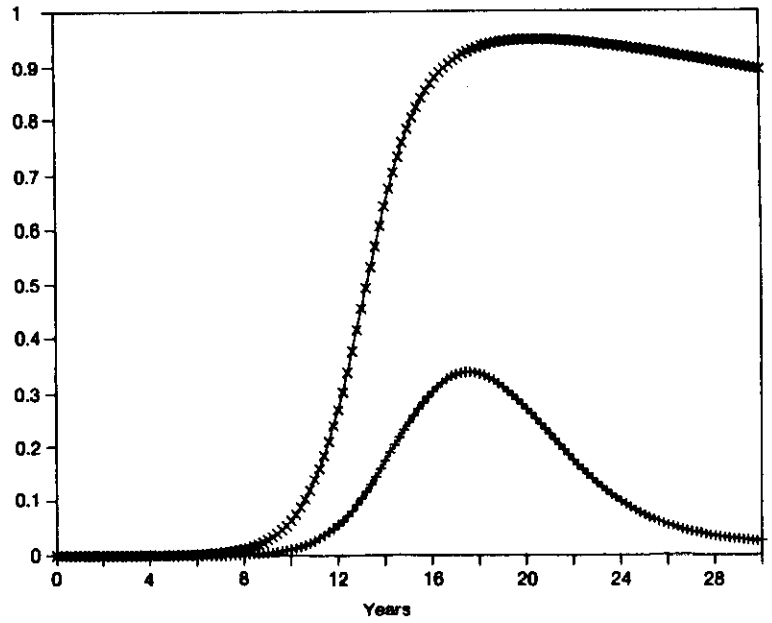

FIG. 12. Temporal solution of the homogeneous-mixing model with recruitment and variable incubation ((4.11) in the text). Proportion seropositive to HIV (×) and the incidence of AIDS (incidence/year/10 000) (+) are recorded on the same graph. Parameter values are as defined in Fig. 10 with $\gamma = 0{\cdot}2$ and $\alpha = 0{\cdot}0628$ (chosen to give a mean incubation period of 5 years).

a constant and independent of the duration of infection. One such comparison is presented in Fig. 12, where the parameter of the function $v(\tau)$ was chosen to give a mean incubation period of five years. Note that, for a fixed value of R_0, variable incubation does not greatly influence the time to maximum incidence (compare with Fig. 10 for the simpler model with constant v). Its main effect is to alter the shape of the epidemic curve in such a way that the rise in cases of 'full-blown' AIDS follows the patterns set by the rise in seropositivity (i.e. incidence of infection), but with a fairly pronounced delay. This contrasts with the simpler model with constant $v(1/v = 4\text{–}5$ years), in which there was a less marked lag between the rise in seropositivity and the rise in AIDS cases. In summary, therefore, the inclusion of an incubation period that is itself sharply defined produces an epidemic curve for AIDS cases that tends initially to track the seropositivity curve with a marked delay, and that tends to have a relatively high proportion of the total number of AIDS cases centred around the peak of the curve.

5. Models with heterogeneity in sexual activity

We now turn to consider an important complication, namely, the transmission of HIV in a population in which sexual activity (defined as the number of different partners per unit of time) varies greatly among individuals (such that the variance in activity greatly exceeds the mean level).

5.1 *The Early Stages of the Epidemic*

As for the homogeneous mixing model, we first examine the early stages of the epidemic to assess how heterogeneity in sexual activity influences the estimation of the basic reproductive rate R_0. The simplest approach to this problem is to divide the total at-risk population of N individuals into subgroups of N_i individuals who have an average of i sexual partners per unit of time: $N_i = NP(i)$ where $P(i)$ is the proportion of the population in the ith class. The number of susceptible and infectious individuals in the ith class at time t are defined as $X_i(t)$ and $Y_i(t)$ respectively. The rate of acquiring infection is taken to be i (number of partners) times λ (probability that a randomly chosen partner infects the susceptible person). Equations (4.1) generalize to

$$\frac{dX_i}{dt} = -i\lambda X_i, \qquad \frac{dY_i}{dt} = i\lambda X_i - vY_i. \tag{5.1a,b}$$

We assume, in the first instance, that v is constant and independent of the duration of infection. The infection probability λ depends on β and on the probability that a partner is infectious. Weighting partners by their degree of sexual activity (i.e. proportionate mixing), we have

$$\lambda(t) = \beta \sum_i iY_i(t) \Big/ \sum_i iN_i(t). \tag{5.2}$$

Equation (5.1a) can be integrated directly to give (R. M. May *et al.*, unpub.)

$$X_i(t) = N_i \exp[-i\psi(t)], \tag{5.3}$$

where

$$\psi(t) = \int_0^t \lambda(g)\, dg.$$

Substituting (5.3) into (5.1b), and summing to produce a differential equation for $\lambda(t)$, we arrive at the result

$$\frac{d\lambda}{dt} = \lambda\left(\beta \sum_i i^2 e^{-i\psi} P(i) \Big/ \sum_i iP(i) - v\right). \tag{5.4}$$

In the early phases of the epidemic, $X_i \approx N_i$ and $\psi = 0$, whereupon (5.4) gives

$$\frac{d\lambda}{dt} = \lambda\left[\beta\left(m + \frac{\sigma^2}{m}\right) - v\right].$$

Here, m is the mean number of partners and σ^2 is the variance. We therefore have exponential growth in the number of infected (as mirrored by temporal changes in $\lambda(t)$) with growth rate $\beta c - v$, where the c in (4.3) of the equivalent homogeneous-mixing model is now, in the heterogeneous-mixing model, defined

as

$$c = m + \sigma^2/m. \tag{5.5}$$

Note that the crude estimation procedure for the basic reproductive rate of HIV infection, based on the data for the doubling time of the epidemic (see (4.5)), remains the same, except that c is now as defined in (5.5). In this equation, the mean and variance of the distribution of sexual activity can be defined in terms of the parameters of a suitable probability distribution that mirrors the empirical data (see Figs. 6, 7, and 8). For instance, the use of a gamma distribution for $P(i)$ permits a good deal of analytical insight (R. M. May *et al.*, unpub.).

The model can be further generalized to take account of the likelihood that, upon acquiring HIV infection, different individuals fall into different classes of subsequent epidemiological histories (or the presentation of different symptoms of disease resulting from infection). Some may remain infectious for relatively short periods before exhibiting 'fully fledged' AIDS or AIDS-related complications (ARC), while others may remain infectious (or non-infectious) for long periods of time without suffering symptoms of disease (the non-AIDS seropositive individuals).

Such a range of possibilities, perhaps based on the mode of acquisition of infection, the viral strain that first infects an individual, or biological characteristics of the individual, can be formally incorporated in the models by letting there be K different kinds of infectious individuals. Equation (5.1a) remains the same but (5.1b) becomes

$$dY_{i,k}(t)/dt = f_k i\lambda(t)X_i - v_k Y_{i,k}.$$

Here, $Y_{i,k}$ is the number of infectious individuals of the infection class k ($k = 1, \ldots, K$) in the ith class of sexual activity; f_k is the proportion of infections that pass to the kth infectious class and v_k is the rate of movement out of this infectious class. The different infectious classes may transmit infection with different efficiencies (per unit of time) so that (5.2) should be generalized to

$$\lambda = \sum_k \beta_k \left(\sum_i iY_{i,k} \right) \Big/ \sum_i iN_i.$$

The numerical analysis of such more general models is relatively straight-forward (given data on the distributions of sexual activity and infectious classes), but the main problem is the proliferation of parameters. Exploration of the properties of this kind of model must await the acquisition of more detailed epidemiological data.

5.2 *The Epidemic in a Population with Recruitment*

To describe the full course of the initial epidemic following invasion we require a more complicated model than that outlined in Section 5.1. For mathematical convenience we employ a continuous framework for the definition of variation in sexual activity (in Section 5.1, a discrete framework was employed (Cox & Anderson, in preparation)). We define $X(t, s)$, $Y(t, s)$, and $I(t, s)$ as the number of susceptibles, the number of infectious people who develop AIDS, and the number of infectious people who do not develop AIDS at time t with sexual activity s (the number of different partners per unit of time), respectively. It is assumed that each individual maintains the same activity status as a susceptible and as an infectious person. We again assume that a proportion p of infectious persons proceeds to 'fully fledged' AIDS and a proportion $1-p$ moves to a non-infectious seropositive state (see Fig. 9b). The model is of the form:

$$\frac{\partial X}{\partial t}(t, s) = \Lambda(s) - sX(t, s)\lambda(t) - \mu X(t, s), \tag{5.6a}$$

$$\frac{\partial Y}{\partial t}(t, s) = psX(t, s)\lambda(t) - (v + \mu)Y(t, s), \tag{5.6b}$$

$$\frac{dI}{dt}(t, s) = (1-p)sX(t, s)\lambda(t) - (v + \mu)I, \tag{5.6c}$$

$$\frac{dA}{dt} = v\int_0^\infty Y(t, s)\,ds - (d + \mu)A, \qquad \frac{dZ}{dt} = v\int_0^\infty I(t, s)\,ds - \mu Z, \tag{5.6d,e}$$

where

$$\lambda(t) = \beta \int_0^\infty s[Y(t, s) + I(t, s)]\,ds \Big/ \int_0^\infty sN(t, s)\,ds. \tag{5.7}$$

Here $A(t)$, $Z(t)$, μ, v, γ, and d are as defined in Section 4.2 (see Fig. 9b). The term $\Lambda(s)$ denotes recruitment to the homosexual population of individuals with sexual activity s, and $N(t, s)$ denotes the total number of persons with activity s.

From (5.6a), the number of susceptibles $x(t, s)$ is

$$X(t, s) = X_0(s) \exp\left(-\mu t - \int_0^t s\lambda(u)\,du\right) + \Lambda(s)\int_0^t \exp\left(-\mu(t-\tau) - \int_\tau^t s\lambda(u)\,du\right) d\tau, \tag{5.8}$$

where $X_0(s)$ is the number of susceptibles of activity s at time $t = 0$. If we define $\bar{Y}(t)$ as the total number of infectious individuals at time t, where

$$\bar{Y}(t) = \int_0^\infty [Y(t, s) + I(t, s)]\,ds, \tag{5.9}$$

then (5.6b) can be reduced to a differential equation for $\bar{Y}(t)$:

$$\frac{d\bar{Y}}{dt} = \lambda(t)\int_0^\infty sX(t, s)\,ds - (v + \mu)\bar{Y}(t). \tag{5.10}$$

If we assume that the distribution of sexual activity is gamma in form with mean m and variance m^2/θ, then (5.10) can be expressed in terms of the parameters of the distribution (m and θ).

The properties of this can be explored using numerical procedures. The scale of the computational problem, however, can be reduced somewhat by employing a

discrete approximation to the continuous gamma distribution of the number of sexual partners per unit of time. A series of sexual-partner classes (of total n) can be defined (e.g. those who have 0–1, 1–5, 5–10, 10–50, 50–100, and 100 plus partners yr^{-1}), and the mean number s_i of partners for the ith class and the mean recruitment rate Λ_i of susceptibles into this class may then be calculated from gamma distributions with defined means (over all classes) and variances. The equations describing the dynamics within each sexual-partner class (with subscript i) are as defined in (4.11) with Λ and c replaced by Λ_i and s_i, respectively. The rate of infection for the discrete sexual-partner class appproximation of (5.6)–(5.7) is defined as

$$\lambda = \beta \sum_{i=1}^{n} s_i[Y(t, i) + I(t, i)] \Big/ \sum_{i=1}^{n} s_i N(t, i).$$

One advantage of this approach to the numerical solution of equations (5.6) to (5.10) is that the discrete approximation to the gamma distribution more accurately reflects the data that are currently available on the sexual habits of homosexuals (see Figs. 7 and 8).

Numerical studies of the discrete sexual-partner class model provide some insight into the manner in which the degree of heterogeneity in sexual activity amongst homosexuals influences the form of the epidemic. Predictions of the influence of changes in the variance-to-mean ratio (m/θ) of the distribution of sexual activity on the incidence of AIDS and the proportion seropositive in the population $(y(t) = [Y(t) + A(t) + Z(t)]/N(t)$ where $N(t) = \int_0^\infty N(t, s)\, ds)$ are presented in Fig. 13.

The graph shows clearly that increasing degrees of heterogeneity in sexual activity decrease the magnitude of the epidemic of AIDS and reduce the maximum level of seropositivity for HIV infection attained during the course of the epidemic. A very high variance-to-mean ratio implies that the population contains a small fraction of highly sexually active individuals who are removed rapidly from the infectious pool (either by death or by passage into the non-infectious seropositive class). As such, the rapid removal of these individuals reduces the total magnitude of the epidemic. The value of the variance-to-mean ratio has little influence on the time period that elapses from invasion to the peak in AIDS cases. The general point illustrated by these numerical simulations is the important influence of heterogeneity in sexual activity on the pattern and magnitude of the epidemic.

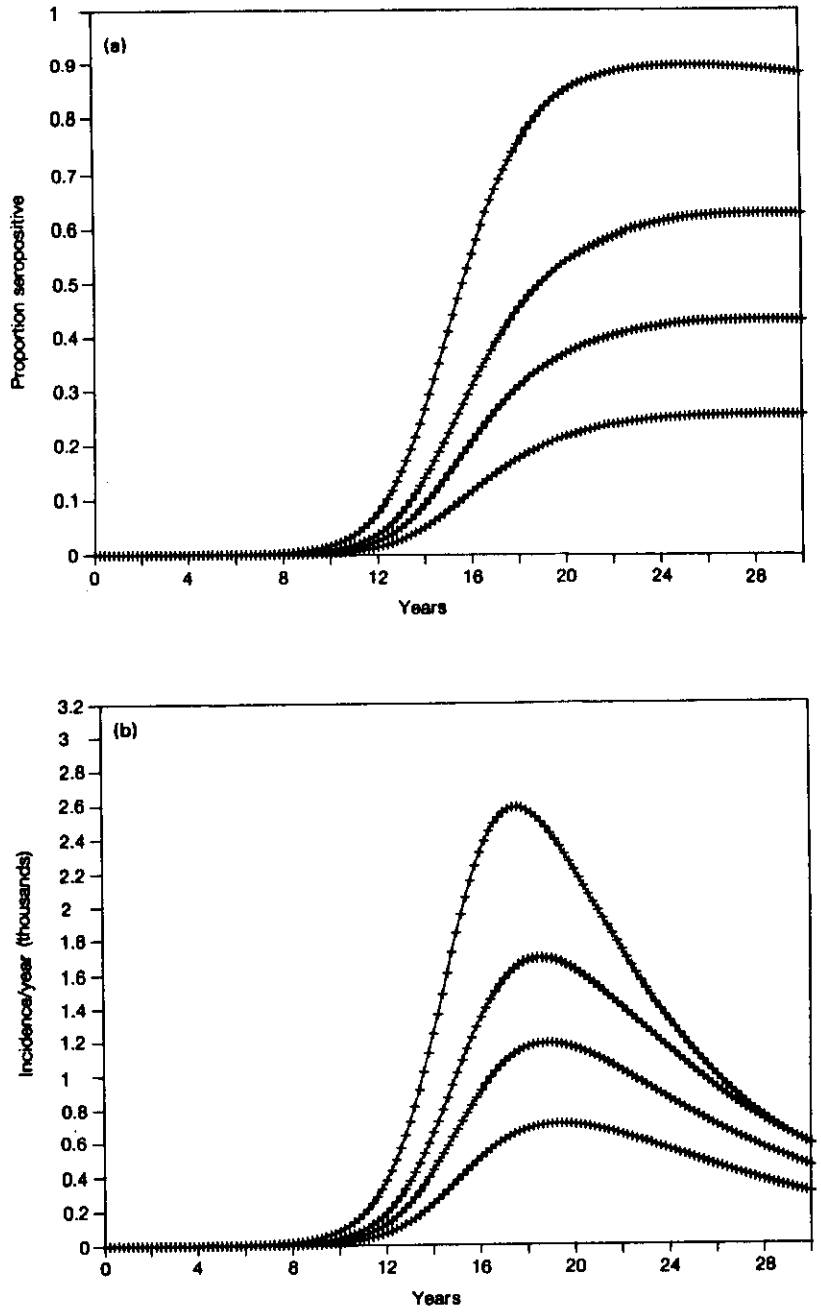

FIG. 13. Temporal solutions of an approximation to the heterogeneous-mixing model with recruitment of susceptibles ((5.6) to (5.10) plus explanation in the main text). The graphs show (a) proportion seropositive to HIV and (b) incidence of AIDS per year for four different variance to mean ratios (m/θ) of the gamma distribution of sexual activity: 1 (top line), 5, 10, 20 (bottom line). Parameter values: $R_0 = 5$, $D = 5$ yrs, $v = 0{\cdot}2\ yr^{-1}$, $d = 1\ yr^{-1}$, $p = 0{\cdot}3$, $N(0) = 100\,000$, $\mu = 1/32\ yr^{-1}$. Six sexual partner classes were defined in the discrete approximation of the gamma distribution: 0–1, 1–5, 5–10, 10–50, 50–100, and 100+ partners yr^{-1}. The Λ_i and s_i were chosen from gamma distributions with the mean number of partners fixed at $5\ yr^{-1}$ and variances of 5, 25, 50, and 100 (to give the variance to mean ratios defined above).

5.3 *Variable Incubation Periods in Heterogeneous Mixing Models*

The model defined by (5.6)–(5.7) can easily be extended to encompass a variable incubation period of the form defined in (3.1b) and (3.2a) (e.g. $v(\tau) = \alpha\tau$). We define $Y(t, s, \tau)$ as the number of infectious individuals (who go on to develop AIDS) at time t of sexual characteristics s who have been incubating the infection for τ time units with the further definition:

$$\hat{Y}(t, s) = \int_0^t Y(t, s, \tau)\,d\tau, \qquad Y'(t, \tau) = \int_0^\infty Y(t, s, \tau)\,ds, \qquad I(t) = \int_0^\infty I(t, s)\,ds.$$

The new model is

$$\frac{\partial X}{\partial t}(t, s) = \Lambda(s) - sX(t, s)\lambda(t) - \mu X(t, s), \tag{5.11a}$$

$$\frac{\partial Y}{\partial t}(t, s, \tau) + \frac{\partial Y}{\partial \tau}(t, s, \tau) = -[v(\tau) + \mu]Y(t, s, \tau), \tag{5.11b}$$

$$\frac{\partial I}{\partial t}(t, s) = (1 - p)sX(t, s)\lambda(t) - (\gamma + \mu)I(t, s), \tag{5.11c}$$

$$\frac{dA}{dt} = \int_0^t v(\tau)Y'(t, \tau)\,d\tau - (d + \mu)A, \qquad \frac{dZ}{dt} = \gamma I - \mu Z, \tag{5.11d,e}$$

where

$$\lambda(t) = \beta \int_0^\infty s[\hat{Y}(t, s) + I(t, s)]\,ds \Big/ \int_0^\infty sN(t, s)\,ds. \tag{5.12}$$

The boundary condition for (5.11b) is given by

$$Y(t, s, 0) = psX(t, s)\lambda(t). \tag{5.13}$$

The model is complex and numerical work (again based on dividing the distribution of sexual partners into a series of discrete classes) is required to explore its properties. Given the insights generated by the inclusion of variable incubation in the homogeneous-mixing model (4.11), it is hardly surprising that the main effect is to delay the interval between invasion and peak incidence, beyond that predicted by the heterogeneous mixing model with v constant. Also, as we saw earlier, variable incubation makes the epidemic more 'spiky' in character (Fig. 14).

6. Discussion

The structures and assumptions incorporated in the models discussed in this paper remain tentative at present. Much is unknown about the epidemiology of HIV, and refinements and alteration to the basic structure of the models will be required in the future, as further biological and epidemiological data become available. The main reason for formulating and exploring such models is to investigate how the various parameters affect viral persistence and spread, and to

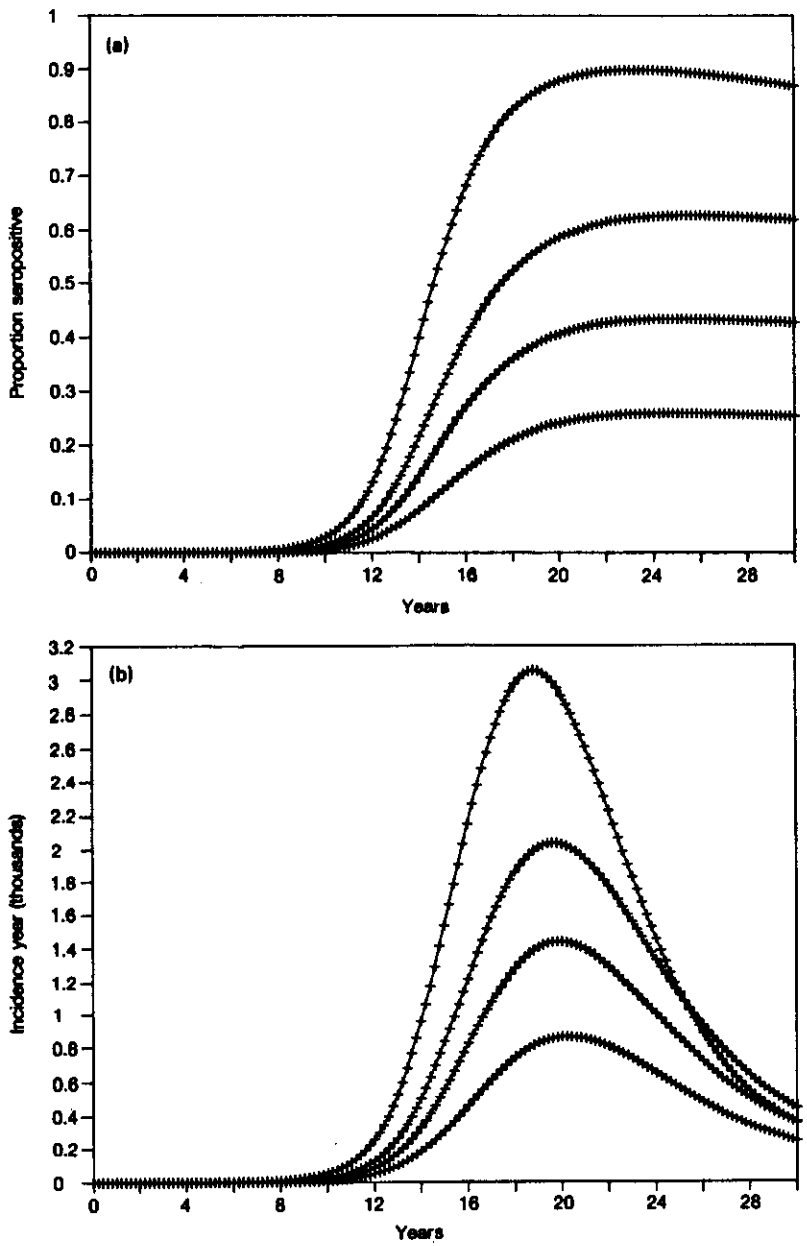

FIG. 14. Temporal solutions of the heterogeneous mixing model with recruitment of susceptibles and variable incubation ((5.11)–(5.13) in the main text). Graphs (a) to (b) as defined in Fig. 13, but with the mean incubation period set at 5 years ($\alpha = 0{\cdot}0628\ \text{yr}^{-1}$; see (3.2) in the text). The numerical methods used to generate these graphs are the same as those used in Fig. 13 and defined in the main

help improve general understanding of the transmission dynamics of HIV infection. We wish to stress that the models are not yet able to generate reliable predictions concerning future trends in AIDS incidence.

The major biological unknowns concern the proportion of those infected with HIV who will progress to develop AIDS, the duration of infectiousness of infected individuals (and whether or not those who proceed to develop AIDS are more or less infectious during the incubation period than those who do not), and the duration of the latent period of infection. At present, reliable data on incubation periods are only available from those infected via blood transfusions. It is possible that such data severely underestimate the average incubation period in those who acquire the virus via sexual contact. Recent work on the rate of appearance of AIDS in HIV seropositive homosexuals (with unknown dates of acquisition of infection) has shown that 34% of individuals had developed AIDS over a three year period (Goedert *et al.*, 1986). This study hints that a much larger proportion of those infected than previously thought may proceed to develop AIDS, and that the average incubation period may be considerably in excess of 4–5 years (Lui *et al.*, 1986). A 10 year average incubation period has been suggested on the assumption that most of those infected with HIV will eventually develop AIDS. If this is the case, and if asymptomatic individuals remain infectious throughout the incubation period, then our crude estimates of R_0 detailed in the first section must be revised upward. For example, if $D = 10$ with a doubling time for cases of infection in the early stages of the epidemic of 9 months, then (4.7) gives an R_0 value of around 10. An illustration of the consequences of this increase is presented in Fig. 15 where the predictions of the homogeneous-mixing model with recruitment, for various values of D and p, are displayed. Broadly speaking, a longer incubation period results in a more drawn-out epidemic (a slower decline in cases following the peak in incidence). Increasing the value of p naturally increases the magnitude of the epidemic. This simple example well illustrates the problems involved in predicting the future course of the current epidemic of HIV given the uncertainties surrounding the biology and epidemiology of the virus.

The heterogeneous-mixing models demonstrate the importance of variability in sexual activity as a determinant of the pattern of the epidemic. High variability results in fewer cases and a lower proportion seropositive to HIV antigens. A substantial fraction of the at high-risk population can remain uninfected, even for moderately large R_0, if variability in sexual habits is sufficiently high. Active people are essentially all quickly infected and hence relatively rapidly removed as infectors before many of the less active individuals are infected. The very simple heterogeneous model for the early stages of the epidemic provides a useful illustration of how the mean and variance of sexual activity combine to determine the magnitude of the force of infection within a community (see (5.5)). Most importantly, this particular model indicates how heterogeneity can influence the measurement and interpretation of key epidemiological parameters.For example, the magnitude of R_0 may be estimated from a knowledge of t_d and D. However, a rough guide to the magnitude of the average transmission coefficient (the chance that an infected partner passes on the infection) can only be obtained via a

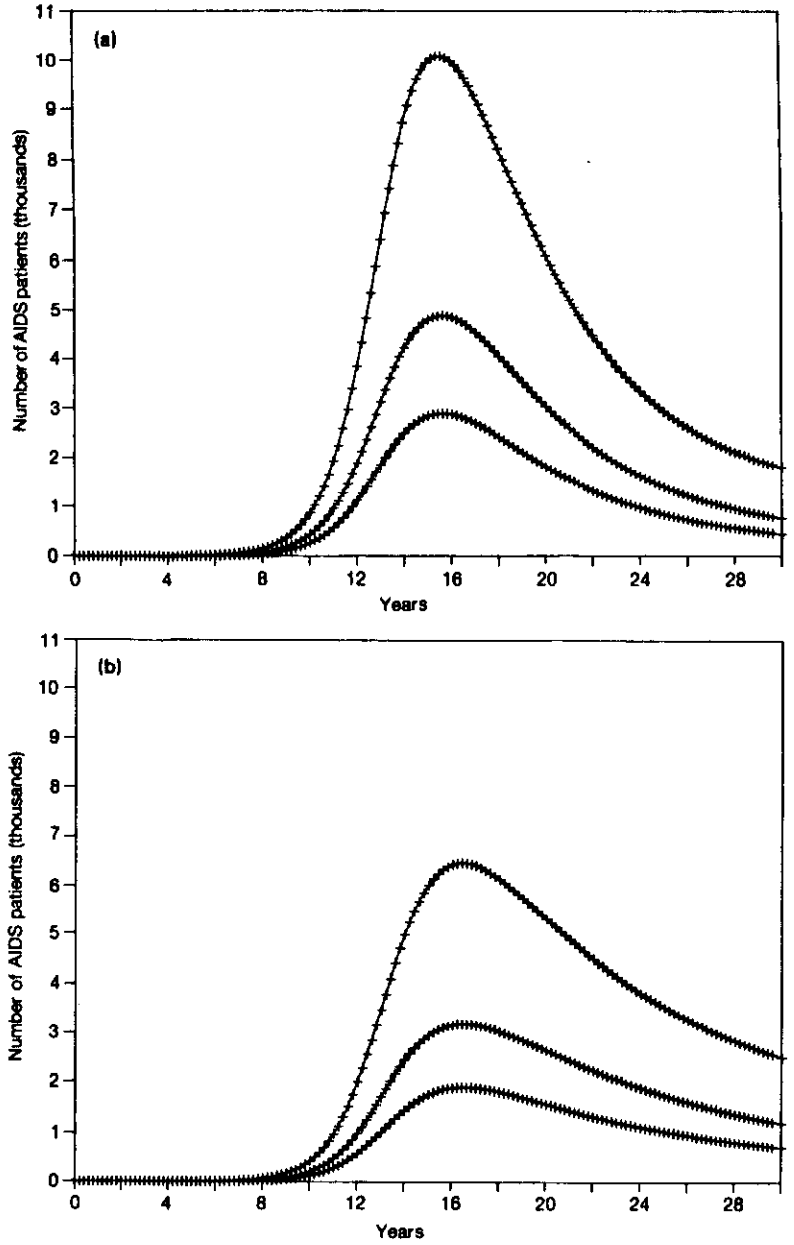

FIG. 15. Temporal solutions of the homogeneous-mixing model with recruitment ((4.9) in the text) showing the effect of varying the incubation (= infectious) period D and of varying the proportion p of those infected who develop AIDS on the number of AIDS patients throughout the epidemic. Parameter values as defined for Fig. 10 but with $D = 5$ yrs and $R_0 = 5{\cdot}62$ in graph (a), and $D = 10$ yrs

knowledge of the mean and variance of the distribution of sexual activity ((4.6), (4.7), (5.5)). Any assessment of how much sexual habits must change in order to reduce R_0 for HIV to less than unity depends not simply on changes in mean activity but also on changes in the variance. A substantial reduction in both could be achieved by limiting or changing the activities of the small proportion of highly promiscuous homosexuals alone, irrespective of changes in the habits of the majority of the population. For example, the unpublished data of Carne & Weller and McManus (Figs. 7 and 8) suggest that the magnitude of the mean level of activity is very much dominated by the activities of a small fraction of very highly active individuals (with 400 or more partners per year).

Our preliminary analyses have centred on the homosexual community and we have ignored other high-risk groups such as IV drug abusers and transfusion recipients. The passage of the infection from homosexuals through bisexual men to heterosexuals, and then between heterosexuals, implies that studies of the overall transmission dynamics of HIV should ideally be based on a matrix of transmission rates between the various risk groups in the total population. For any two groups, the passage of infection between them is likely to be asymmetrical, due, in part, to unequal population sizes. Data of sufficient quantity to define such a matrix of transmission coefficients are unavailable at present, but their collection should clearly be a priority in order to assess the likely spread of the virus through the heterosexual community. Transmission between heterosexuals is known to occur (via studies of the partners of heterosexuals infected by blood transfusion), and it appears to be an efficient route for the spread of the virus (Vogt *et al.*, 1985; Des Jarlais *et al.*, 1984). In certain parts of Africa, heterosexuals appear to be the major risk group. This may be associated with differences in sexual and other practices that facilitate virus transmission (Biggar *et al.*, 1986; Piot *et al.*, 1984). To what extent the virus will be able to persist within a wide cross-section of adult society in Europe and North America remains to be seen.

However, it is possible that, following its recent establishment in heterosexual population, HIV could spread rapidly and persist endemically. Whether the virus can maintain itself within heterosexual communities depends ultimately on whether the values of β (transmission probability) and c (effective average number of sexual partners)—although lower than the corresponding values for transmission among homosexual males—are nevertheless large enough to keep R_0 above unity. As far as the homosexual community is concerned, the prevailing variability in sexual activity within a population will have a major influence on the possible rate of spread of HIV, on the total magnitude of the initial epidemic, and on the equilibrium endemic prevalence of infection.

Finally, it is worth considering briefly the future role of mathematical models in the study of AIDS. An immediate problem that springs to mind is: 'why bother with simple models, when those of great complexity, perhaps encapturing transmission between many at-risk groups, can be explored by numerical means.' At this stage of data collection and given current biological and epidemiological knowledge, it seems sensible to concentrate on models that focus on what information can and cannot be extracted from the sparse data available.

Estimates of the doubling time can be made, and they yield crude measures of the overall basic reproductive rate within given at-risk groups (in particular, the dominant one—homosexual males). The information is not sufficient as yet to determine how this overall rate is made up of contributions arising from cross-transmission between risk groups. One hopes it will become so, as cohort studies progress. Data on the peak incidence of AIDS in the initial epidemic within a particular at-risk group will tell us about the proportion of infecteds who go on to develop AIDS, the duration of the incubation period, and variability in sexual activity. Similarly, the maximum prevalence of seropositivity to HIV antigens will provide information on R_0 and variability in sexual activity (May and Anderson, 1987).

As the intensity of epidemiological research increases, simple models will be of less value once the various unknowns and complications that have been hinted at in this paper are defined and quantified. When this stage is reached, such models should provide tools for predictive work to help assess future health-care and service needs. The last year has seen increasing pressure on government and health services in the U.K. to take preventative action to control the HIV epidemic and to encourage behavioural changes. Models may help in determining the magnitude and type of behavioural change which has to be achieved to control the spread of infection.

Acknowledgements

We thank the M.R.C. for financial support of part of this research (to R. M. A.) and Drs. I. Weller, C. Carne, and T. MacManus for access to unpublished data on the sexual activity of homosexuals in London.

This article is based on a paper read at the IMA Conference on 'The Mathematical Theory of Biological Systems' held in Oxford, 7–9 July, 1986.

References

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

Anderson, R. M., & May, R. M. 1986 The invasion, persistence and spread of infectious diseases within animal and plant communities. *Phil. Trans. R. Soc. Lond. B.* **134**, 533–570.

Barre-Sinioussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. 1983 Isolation of a *T*-lymphotropic retrovirus from a patient at risk for acquired immune deficiency (AIDS) *Science* **220**, 868–871.

Biggar, R. J. 1986 The AIDS problem in Africa. *The Lancet* I, 79–83.

Carne, C. A., Weller, I. V. D., Sutherland, S., Cheinsong-Popov, R., Ferns, R. B., Williams, P., Mindel, A., Tedder, R., & Adler, M. S. 1985 Rising prevalence of human *T*-lymphotropic virus type III (HTLV-III) infection in homosexual men in London. *The Lancet* I, 1261–1262.

C.D.C. 1982 *M.M.W.R.* **31**, 507–514.

C.D.C. 1985a *M.M.W.R.* **34**, 573–582.

C.D.C. 1985b *M.M.W.R.* **34,** 583–589.
C.D.C. 1985c *M.M.W.R.* **34,** 681–695.
C.D.C. 1986 *M.M.W.R.* **35,** 17–21.
Coffin, J., Hasse, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., & Weiss, R. 1986 Human Immunodeficiency Viruses. *Science* **232,** 697.
Cooke, K. L., & Yorke, J. A. 1973 Some equations modelling growth processes and gonorrhoea epidemics. *Math. Biosci.* **16,** 75–101.
Cox, D. R., & Oakes, D. 1984 Analysis of survival data. Chapman & Hall, London.
Des Jarlais, D. C., Chamberland, M. E., Yancovitz, S. R., Weinberg, P., & Friedman, S. R. 1984 Heterosexual partners: a large risk group for AIDS. *The Lancet* II, 1346–1437.
Esteban, J. I., Shih, J. W. K., Chang-Chih Tai, Bodner, A. J., Kay, J. W. D., & Alter, H. J. 1985 Importance of western-blot analysis in predicting infectivity of anti-HTLV-III/LAV positive blood. *The Lancet* II, 1083–1086.
Gallo, R. C., Salahuddin, S. Z., Popovic, M., Sheaver, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., & Markham, P. D. 1983 Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224,** 500–503.
Goedert, J. J., Sarngadaran,M. G., Biggar, R. J., Weiss, S. H., Winn, D. M., Grossman, R. J., Greene, M. H., Bodner, A. J., Strong, D. M., Gallo, R. C., & Blattner, W. A. 1984 Determinants of retrovirus (HTLV-III) antibody and immunodeficiency conditions in homosexual men. *The Lancet* II, 711–715.
Goedert, J. J., Biggar, R. J., Weiss, S. H., Eyster, M. E., Melbye, M., Wilson, S., Ginzburg, H. M., Grossman, R. J., DiGiola, R. A., Sanchez, W. C., Giron, J. A., Ebbeson, P., Gallo, R. C., & Blattner, W. A., 1986 Three year incidence of AIDS in five cohorts of HTLV-III-Infected risk group members. *Science* **231,** 992–995.
Hahn, B. H., Gonda, M. A., Shaw, G. M., Popovic, M., Hoxie, J. A., Gallo, R. C., & Wong-Staal, F. 1985 Genomic diversity of the acquired immune deficiency syndrome virus HTLV-III: different viruses exhibit greatest divergence in their envelope genes. *Proc. Nat. Acad. Sci.* **82,** 4813–4817.
Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S., & Parks, W. P. 1986 Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232,** 1548–1553.
Hethcote, H. W., & York, J. A. 1984 Gonorrhoea: transmission dynamics and control. *Lecture Notes in Biomathematics* **56,** 1–105.
L'Age-Stehr, J., Schwarz, A., Otterman, G., Langmaack, H., Bennhold, I., Niedrig, M., & Koch, M. A. 1985 HTLV-III infection in kidney transplant recipients. *The Lancet* II, 1361.
Lui, K. J., Lawrence, D. N., Morgan, W. M., Peterman, T. A., Haverkos, H. W., & Bregman, D. J. 1986 A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proc. Nat. Acad. Sci.* **83,** 3051–3055.
May, R. M., & Anderson, R. M. 1987 The transmission dynamics of HIV infection. *Nature* (in press).
McKusick, M. L., Horstman, W., & Coates, T. J. 1985 AIDS and sexual behaviour reported by gay men in San Francisco. *Am. J. Public Health.* **75,** 493–496.
Marasca, G., & McEvoy, M. 1986 Length of survival of patients with acquired immune deficiency syndrome in the United Kingdom. *BMJ* **292,** 1727–1729.
Mortimer, P. P., Vandervelde, E. M., Jesson, W. J., Pereira, M. S., & Burkhardt, F. 1985 HTLV-III antibody in Swiss and English intravenous drug abusers. *The Lancet* II, 449–450.

Moss, A. R., McCallum, J., Volberding, P., Bacchetti, P., & Dritz, S. 1984 Mortality associated with mode of presentation in the acquired immune deficiency syndrome (AIDS). *JNCI* **73,** 1281–1284.
Peterman, T. A., Drotman, D. P., & Curran, J. W. 1985 Epidemiology of the acquired immunodeficiency syndrome (AIDS) *Epidemiologic Reviews* **7,** 1–21.
Peterman, T. A., Jaffe, H. W., Feorino, P. M., Getchell, J. P., Warfield, D., Haverkos, H. W., Stoneburner, R. L., & Curran, J. W. 1985 Transfusion-associated acquired immunodeficiency syndrome in the United States. *J.A.M.A.* **254,** 2913–2917.
Piot, P., Quinn, T., Taelman, H., Feinsod, F. M., Minlangu, K. B., Wobin, O., Mbendi, N., Mazebo, P., Ndangi, K., Stevens, W., Kalambayi, K., Mitchell, S., Bridts, C., & McCormick, J. B. 1984 Acquired immune deficiency syndrome in a heterosexual population in Zaire. *The Lancet* II, 65–69.
Salahuddin, S. Z., Groopman, J. E., Markham, P. D., Sarngadharan, M. G., Redfield, R. R., McLane, M. F., Essex, M., & Sliski, A. 1984 HTLV-III in symptom-free seronegative persons. *The Lancet* II, 1418–1420.
Vogt, M. W., Witt, D. J., Craven, D. E., Byington, R., Crawford, D. F., Schooley, R. T., & Hirsch, M. S. 1986 Isolation of HTLV-III/LAV from cervical secretions of women at risk for AIDS. *The Lancet* I, 525–527.
Weber, J. N., Wadsworth, J., Rogers, L. A., Moshtael, O., Scott, K., McManus, T., Berrie, E., Jefferies, D. J., Harris, J. R. W., & Pinching, A. J. 1986 Three-year prospective study of HTLV-III/LAW infection in homosexual men. *The Lancet* I, 1170–1182.
Weiss, R. A. 1982 The persistence of retroviruses. In: *Virus Persistence* (B. W. J. Mahy, A. C. Minson, & G. K. Darby, Eds). Cambridge University Press, London. pp. 267–288.
Wong-Staal, F., & Gallo, R. C. 1985 Human *T*-lymphotropic retroviruses. *Nature* **317,** 395–407.