Using Mathematical Models to Understand the AIDS Epidemic

JAMES M. HYMAN

AND

E. ANN STANLEY

Center for Nonlinear Studies, Theoretical Division, MS-B284,
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received 1 June 1987; revised 2 February 1988

ABSTRACT

The most urgent public-health problem today is to devise effective strategies to minimize the destruction caused by the AIDS epidemic. This complex problem will involve medical advances and new public-health and education initiatives. Mathematical models based on the underlying transmission mechanisms of the AIDS virus can help the medical/scientific community understand and anticipate its spread in different populations and evaluate the potential effectiveness of different approaches for bringing the epidemic under control. Before we can use models to predict the future, we must carefully test them against the past spread of the infection and for sensitivity to parameter changes. The long and extremely variable incubation period and the low probability of transmitting the AIDS virus in a single contact imply that population structure and variations in infectivity both play an important role in its spread. The population structure is caused by differences between people in numbers of sexual partners and the use of intravenous drugs and because of the way in which people mix among age, ethnic, and social groups. We use a simplified approach to investigate the effects of variation in incubation periods and infectivity specific to the AIDS virus, and we compare a model of random partner choices with a model in which partners both come from similar behavior groups.

I. INTRODUCTION

Most current predictions of the acquired immune deficiency syndrome (AIDS) epidemic are based on simple exponential or polynomial extrapolations of current trends. These curve-fitting methods cannot be used reliably for long periods of time, nor can they provide understanding of the interactions that lead to the epidemic's spread. During the long asymptomatic period after infection with the human immunodeficiency virus (HIV) that

causes AIDS, changes in the environment of viral transmission occur continuously, causing complex interactions. Only models that are founded on the transmission mechanisms of HIV can show how the early infection of high-risk groups, behavioral changes, and future medical advances such as treatments and vaccines will affect the future course of this epidemic. The effects will be highly nonlinear functions of the parameter values and at times may even lead to changes that are counter to both intuition and simple extrapolated predictions. The mathematical model predictions of these counterintuitive mechanisms may greatly improve our understanding of the observations.

In developing the mathematical models, we are creating a logical structure that organizes existing information on AIDS into a coherent framework and suggests new information that must be collected about a wide variety of topics, such as drug use, sexual activity, and the interactions between HIV and the immune system. Models can provide qualitative insights, even when data are lacking, and can help prioritize data collection.

We have already gained some qualitative insights from our modeling work. For example, we have seen that the amount of sexual contact and needle-sharing between a small group of high-activity and a larger group of lower-activity individuals determines both who gets infected and the speed with which the epidemic progresses. If there is little mixing between these groups, then the individuals in high-risk groups are nearly all infected before the infection moves into lower-risk groups. However, if mixing is large, many more lower-risk individuals will be infected in the early stages of the epidemic; the epidemic moves much faster when mixing is large because there is a larger pool of lower-risk individuals to feed it. In a model where partners are chosen randomly, regardless of their partner-change rate, the total number of infected low-risk individuals quickly exceeds the number of infected high-risk individuals. This random-mixing result is inconsistent with the data. These differences support the urgent need to collect and analyze the information on mixing patterns to estimate critical model parameters.

The probability of infection per contact (infectivity) is too poorly understood to use the AIDS caseload data to distinguish between these mixing patterns. Our modeling also indicates that, if the difference between male-to-female and female-to-male infectivity is large, then the lower of these two infectivities will tend to determine heterosexual spread, with epidemic patterns potentially different from that seen in homosexuals and intravenous (IV-drug) users. This difference indicates that collecting and analyzing information on infectivity should have high priority.

The infectivity is significantly lower in the middle stage of HIV infection than when end-stage disease (AIDS) approaches. Therefore, testing and counseling programs that identify and persuade infected individuals to avoid

infecting others will be more effective than if the infectivity were constant. Models to predict the role of testing and counseling must include the effects of variable infectivity.

Although it is unlikely that any one model will provide accurate long-term predictions of the numbers of AIDS cases, a model that is based on interactions that lead to disease transmission could eventually allow investigators to answer many questions. For example, one could assume increased condom use by people in a targeted age group and region and then determine how much that increased use will slow the local course of the epidemic. This predictive ability would then help authorities decide if it is more effective to encourage condom use in that group than to use another strategy, such as stressing the importance of having fewer partners or reducing the incidence of other sexually transmitted diseases (STDs), to lower the probability of infection for some population groups. As another example, a partially effective vaccine with potentially harmful side effects might be developed. Somehow it must be ascertained which persons should be vaccinated. The model would be used to understand how vaccinating each group affects the spread of the epidemic.

To prevent new infections, intervention strategies must focus on the groups currently being infected, and those next at risk. Although the most accessible and dramatic data come from AIDS cases, these cases primarily represent infections that occurred 4 or more years ago. To understand where infections are occurring today is a difficult task. Models can help in planning future seroprevalence studies and intervention strategies by indicating where the epidemic front lines are likely to be.

As models are developed, they must be tested for consistency with the past history of the epidemic. We cannot hope to predict the future before we can explain the past. Much of the focus of this paper is therefore on understanding past HIV spread in homosexual men.

Any inconsistencies between the data and the models need an explanation: matching parameters so that the absolute numbers of AIDS cases are correct is not a verification that a model is correct. Many different models can match these gross data sets and forecast widely different futures. Parameter estimates must lie within ranges obtained by independent observations. Correlated residuals between the fitted model predictions and AIDS data may give important clues to additional mechanisms that models must incorporate. Data from seroprevalence and cohort studies should also be consistent with the model's predictions. For example, a random-mixing model leads to a fast early growth in infection in homosexual men having 2–5 partners per year. This growth rate is inconsistent with the data from testing blood samples obtained before 1982 [12, 22] and also with the Center for Disease Control (CDC) case-tracing study of the first men with AIDS [6]. On the other hand, in a model where high-risk individuals primarily mix

with others at high risk, then lower-risk groups are not infected in the early stages. This model is consistent with the AIDS data and agrees with the seropositivity studies. We plan to test the hypothesis that most mixing was between men of similar risk behavior by analyzing the San Francisco Hepatitis B data on behavior versus infection from 1978 to 1982.

Another use of models is to estimate unknown data on the basis of the known facts. For example, the past distribution of HIV infection can be estimated from the current AIDS caseload and the distribution of times from infection to AIDS. To determine the consistency of the generated data requires a formal mathematical model similar to the one we are designing. The available data can also be assessed indirectly to determine their internal consistency by leaving some data out, generating estimates of the missing data based on one or more models, and then comparing the two data sets.

The HIV that causes AIDS is primarily transmitted through sexual contact (man-woman, man-man), sharing of hypodermic needles, and exposure to infected blood either perinatally or through blood transfusions. HIV is not transmitted by nonsexual daily contacts, even though the virus has been isolated from almost every body fluid [17]. The infection risk to an individual depends both on the behavior of the individual and on the prevalence of infection in the groups with which the individual has sexual contacts or shares needles. This prevalence varies between regions and age groups, as well as between behavioral risk groups. An individual is more likely to become infected if he or she has multiple sexual partners; has sexual partners in a high-risk group; lives in a highly populated area; lives in the New York City, Washington (D.C.), San Francisco, or Los Angeles area; shares needles when using drugs; is between 25 and 35 years of age; or has another STD.

A single model that tried to address all of the questions raised in this paper would contain too many variables to be solved numerically on even the largest and most advanced computers. Even if it were possible to solve the system, not enough is known about human behavior to supply the necessary information to the program, nor would a deep understanding of the interactions within the transmission network be gained by initially solving a large system. Instead, simplified submodels must be developed to address specific questions. The assumptions behind these models should be clear, including both what is being neglected that can probably be neglected and what is being neglected that is unrealistic. Studying families of simple models will allow us to understand how different factors interact in the spreading of the AIDS virus.

For example, to comprehend how precisely the infectivity profile (infectiousness with time since infection) must be measured, one can look at the sensitivity of a very simple model to variations in the profile. Such a model can lump age groups and regions, but it cannot ignore all heterogeneities in

sexual-partner choices. On the other hand, if we wish to understand how age differences may delay spreading of the infection from one age group into another, then we cannot ignore age-structured behavior. The behavior of simple models should be carefully investigated to build a picture of interactions that will allow us to make estimates that lead to simplifications in more global models.

For modeling purposes, the portion of the male and female population that engage in behaviors that put them at risk for HIV, namely, nonmonogomous sexual contact and needle-sharing drug use, is divided according to their risk behaviors and the manner in which they choose partners. These susceptible people are infected through contact with infected people, and infected people develop clinical AIDS (such as Kaposi's sarcoma [KS] or opportunistic infections such as pneumocystis pneumonia [PCP]) at a rate that depends on the length of time since HIV infection. AIDS patients subsequently die at a rate that depends on the length of time since they developed AIDS and on the type of clinical manifestation (either KS or opportunistic infections). We assume that an infected person remains infected and infectious for life. This one-way migration of susceptibles to infecteds is due to the chromosomal integration of the proviral DNA into the host cell.

In the next three sections, we discuss many of the risk factors and aspects of the AIDS virus that we foresee as being important to the epidemic and some that will eventually be found to be unimportant. The future spread of the virus in the United States and Europe will most likely be through sexual contact and drug needle sharing. A model of the transmission pattern in Africa would require also including blood transfusions and perhaps other factors.

In Section IV, we discuss the growth of AIDS cases in the United States to date. The total number of cases has grown as time cubed, within a few percent. We use an extrapolation of this cubic and estimates of the distribution of times from infection to AIDS diagnosis to estimate the growth in the number infected.

In Section V, we present simple models, which are chosen to allow investigation of a particular set of questions about the epidemic that has occurred so far in the United States. These questions include the sensitivity of models to the variation in infectiousness as time since infection, the effect that random or biased partner choice has on the shape of the epidemic, and the importance of multiple contacts between partners. In Sections VI and VII we discuss parameter estimates and present numerical investigations of these models.

As we discuss the issues that are important for modelers to consider, we will be providing a logical structure for the diverse data that researchers are collecting. Also, new questions and insights will arise to guide investigators

in directing their research to add to the general understanding of this epidemic.

II. POPULATION RISK STRUCTURE

In contrast to our current understanding of the transmission of malaria [1, 50], measles [15], rubella [4], rabies [44], and many other diseases (Anderson and May, [5]), little is known about modeling the behavior of STDs in the sexually active community. To analyze the HIV transmission dynamics, the sexual activity and needle-sharing drug use of the susceptible population must first be understood and modeled. These activities, about which little is known, pose formidable research questions in themselves.

The risk group from which a person chooses partners for sex or needle sharing is an important social question about which little is known. The married man who has an affair with a married woman takes a different risk from one who solicits a prostitute once a year. Both men may have the same number of new partners each year, but they have chosen those partners in a very different manner.

Risk also depends on the infectiousness of each contact, which depends on the type of contact, the use of protective measures, and how far along the infected person is in the course of the infection. It is perhaps important to note that the infectiousness of HIV is sufficiently low that the spouse of an infected person may not become infected until about a year before AIDS develops [17, 21], so that a person does not necessarily become infected if his/her long-term partner does.

Some recent information on the amount and type of drug abuse in the United States is available from the National Survey on Drug Abuse conducted by the National Institute on Drug Abuse. On the other hand, no large-scale studies specifically aimed at sexual behavior have been conducted in the United States since the Kinsey studies more than 35 years ago. However, a number of other studies, such as fertility studies, have included some questions on sexual behavior or have studied specific groups. Several ongoing efforts involve searching through these studies for information relevant to HIV spread (John Gagnon at SUNY at Stony Brook, Wendy Cain at the National Institute for Child Health and Human Development [NICHD]). In addition, NICHD is designing and will implement a nationwide survey of sexual behavior and needle-sharing behavior specifically aimed at gathering information about the transmission of the AIDS virus.

Endemicity of the infection also plays a major role. Once the infection becomes endemic in a group of people, it may spread in that group fairly rapidly, whereas another group that has few contacts with infected groups may remain protected for a long time. Age differences, physical distance, ethnicity, and other social groupings may all provide barriers to the spread-

ing of infection. Behaviors also vary between different groups of people, leading to different spreading rates in different groups.

The risk-group divisions we have identified as being of possible importance to the spreading of this epidemic include the following:

a: age,
r: number of new male partners per year,
s: number of new female partners per year,
g: sexual activity group,
d: number of people with whom needles are shared per year,
p: population density,
z: zone of the country,
e: ethnicity or social group, and
c: cofactors.

Some of these (a, p, z, e) act as barriers to the spread of the disease. That is, people of similar ages and ethnicity who are living in nearby geographic regions are more likely to spread the virus among themselves than they are to other groups. Other factors (r, s, g, d, c) determine how the spreading occurs within social groups.

The transmission of AIDS involves long time scales, and therefore members are not frozen into a given risk group once they have entered it. This flow occurs because behavior changes with age, marital status, knowledge of infection, changing social mores, and educational efforts and because of movement of people between geographic regions. This flow is an additional source of contact between risk groups.

A. AGE

Age is important for a number of reasons. There is a distribution of ages at which people become sexually active and presumably tend to migrate first into more active groups and then into long-term relationships as they age. Drug use is age dependent. The use of particular drugs, such as heroin, goes in and out of style and is thus generational (R. Chaisson and A. Moss, UC San Francisco, personal communication). There are natural barriers to contacts between age groups, so that the infection will not spread between age groups as rapidly as within an age group. Social groups, such as high school or college students, are age dependent. The amount and type of traveling done also are age dependent. The number of children born with the HIV infection will depend on the number of infected women who are having children, which varies with age. Death rates, health, and disease progression [5, 7] are age dependent.

In regions where AIDS becomes a major problem, as it already has in central Africa, this epidemic has the potential to deplete the productive age

groups in the short run and to entirely change the population's age structure in the long run [41].

B. SEXUAL ACTIVITY

Risk from sexual activity depends on the probability of choosing an infected partner as well as on the number and type of contacts with an infected partner. The probability of choosing an infected partner depends not only on how many new partners are chosen but also on the manner in which those partners are chosen. There is a wide variation in the rates that sexually active people and needle-sharers change partners (see Figure 10 in Section VI). A small core of HIV-infected, very sexually active people can drive the epidemic.

Most models for the transmission of venereal diseases [28, 3] have assumed that all partners are picked at random from the pool of available partners. This assumption leads to the proportionate-mixing assumption that the per-year probability of someone with i partners per year picking an infected partner with j partners per year is $i \cdot j \cdot P_j / P_T$, where P_j is the number of infected people with j partners per year and P_T is the total number of partners picked per year. These models also assume that the probability of infection per partner is the same. However, it is clear that these assumptions are simplistic.

In our models, we assume that an average probability of infection can be assigned to each contact. This assumption may not be sufficiently accurate to predict the spread of HIV, and additional factors may need to be included in the model. For example, the probability of infection might depend strongly upon the strain of the virus or on the health of the partners.

There is probably a tendency for people with fewer partners to have more contacts per partner than do people with many partners. There is also a bias of like toward like, so that people with few partners tend to choose partners who also have few partners. Adding these biases into the Anderson et al. model leads to substantially different predictions from their random-mixing model with equal risks.

Another aspect of behavior is that most sexually active people, both homosexual and heterosexual, move in and out of stable partnerships [35]. They may go into the dating pool and have a number of short-term relationships with a small number of contacts per person before forming a new partnership, or they may go directly from one partnership to the next (with or without some overlap). Sexual partnerships have a wide variation in their duration. The duration of each partnership depends on the sexual-activity groups of the partners involved. The more sexually active people in the dating group form shorter partnerships than the less-active individuals. A similar dependence holds for the duration of abstinence periods. The duration of the longer-term partnerships tends to increase with age. A recent

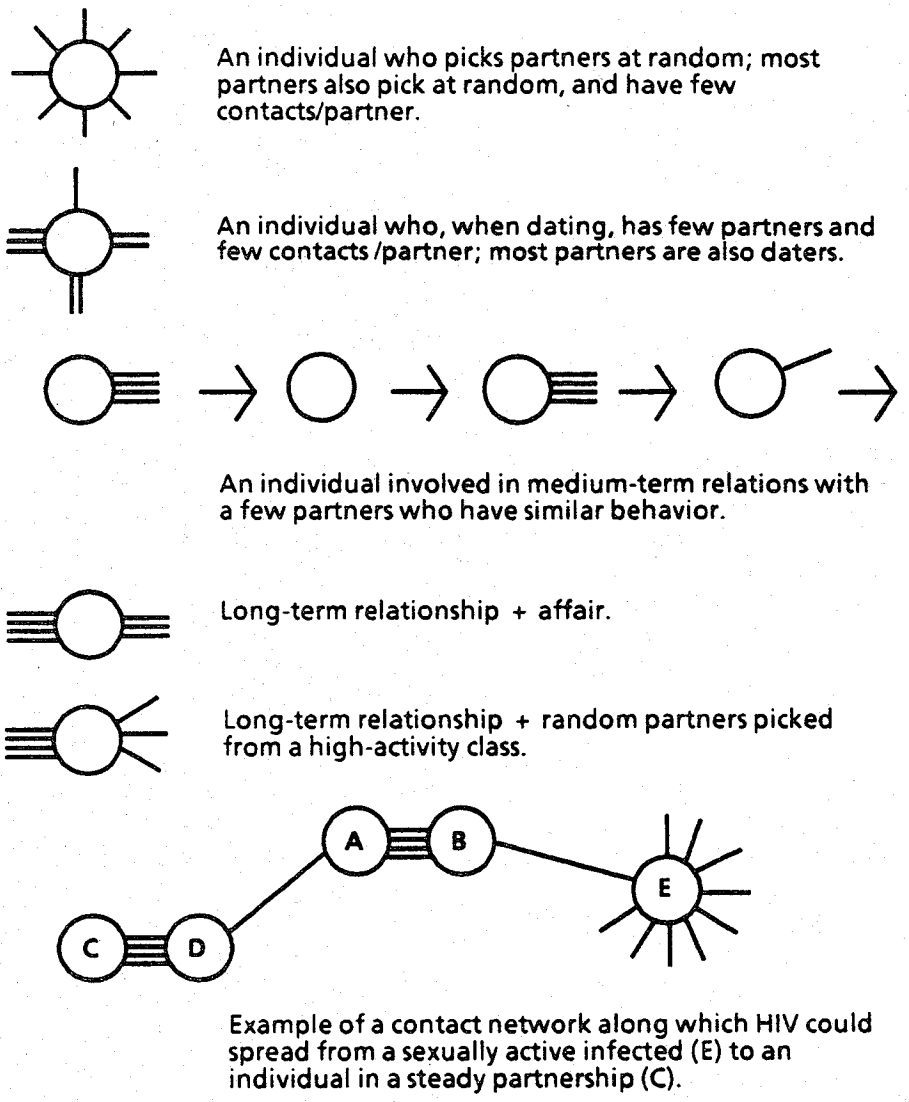


FIG. 1. Different individuals (indicated by circles) may have very different sexual contacts (indicated by the lines).

model for the spread of AIDS by Klaus Dietz [14] incorporates some of these flow ideas using survey data of the West German population.

Also, a fraction of the population maintains long-term relationships and then has a certain number of outside partnerships. The risk to individuals from longer-term relationships depends on the outside partners or the previous partners of their mates. Some possible behavior classes are shown in Figure 1.

Although the data are poor at this point, the infectiousness of a contact may depend on the type of contact (man-man, woman-man, man-woman, anal-genital, oral-genital). Infectiousness also depends on other cofactors such as venereal diseases and the use of protective devices (condoms, nonoxynol-9). We need estimates for the prevalence of these cofactors, how

frequently protective devices are used, and how much behavior can be influenced by factors such as education, knowledge that a partner or oneself is infected, and fear of infection. Also, individuals with higher-risk behavior are more likely to seek testing and discover their infection than are those involved only in low-risk behavior. As public awareness increases and more people know they are infected, we speculate that the resulting drift toward safer sexual practices will slow the spread of the virus.

The infected-spouse studies [17] and the African epidemic demonstrate that the virus can spread through a heterosexual network. Growing evidence suggests that the fast heterosexual spread in Africa is partly due to a high prevalence of cofactors, such as genital ulcers caused by chanchroids, which may greatly increase both infectiousness and susceptability. In the developed world, such severe cofactors are virtually nonexistence. However, other cofactors are present, such as gonorrhea, syphilis, and herpes, that may increase transmission rates less dramatically. Without data on infectiousness, with and without cofactors, male to female and female to male, it is impossible to tell whether or not a self-sustaining heterosexual epidemic will occur in the United States. The few current heterosexual AIDS cases are primarily driven by the epidemic among homosexuals and IV drug users. A slowly growing heterosexual epidemic could be masked by cases due to contacts with these groups. It is unlikely that models can distinguish between these two possibilities without estimates of transmission probabilities from partner studies (e.g., [17, 45]).

Although approximately the same number of men as women are infected with HIV in central Africa, and in some groups of military recruits in the United States [9], this fact does not imply that the virus is transmitted with equal efficiency between men and women, even in the presence of cofactors [40]. The numbers of partners that each has may also play a big role. The infected women may have had, on the average, far fewer partners than the infected men, but there may be a pool of infected prostitutes with whom many men have contact. Also, the presence of other STDs may be a more important cofactor in a heterosexual transmission network than in a homosexual network.

Consider HIV transmission through a simplified heterosexual network, where one male infects one female, who in turn infects another male,

$$M \overset{\beta_r}{\rightarrow} W \overset{\beta_i}{\rightarrow} M \overset{\beta_r}{\rightarrow} W \rightarrow ,$$

and a simplified homosexual network, where each male infects one more and all are assumed to engage with equal frequency in both insertive and receptive anal intercourse,

$$\rightleftarrows M \rightleftarrows M \underset{\alpha_i}{\overset{\alpha_r}{\rightleftarrows}} M \rightleftarrows ,$$

and where the transmission (infectivity) rates are

β_r for man-to-woman (receptive),
β_i for woman-to-man (insertive),
α_r for man-to-man (receptive), and
α_i for man-to-man (insertive).

This heterosexual transmission chain looks like several resistors in series with resistivities β_r^{-1} and β_i^{-1}. For the $M \to W \to M$ chain, the two "resistivities" add, giving an average per-link resistivity of $\frac{1}{2}(\beta_r^{-1} + \beta_i^{-1})$ and thus an average transmission rate per link of $\beta = 2\beta_r\beta_i/(\beta_r + \beta_i)$. For the homosexual chain the transmission routes are in parallel and the average transmission rate per link is $\alpha = \alpha_r + \alpha_i$. A slightly more realistic model, where each person can have more than 2 partners, and for which the average transmission rate is somewhat modified, is discussed in Section V.A.

If, as some have proposed, in the absence of other cofactors such as STDs the probability of being infected during insertive intercourse is much less than in receptive intercourse (that is, $\beta_i \ll \beta_r$, $\alpha_i \ll \alpha_r$), then the average transmission rates would be $\beta \simeq 2\beta_i$ and $\alpha \simeq \alpha_r$. The heterosexual transmission rate would be governed almost entirely by $W \to M$, the insertive infectability, whereas homosexual transmission would be driven by the faster receptive transmission rate. Thus, the most effective strategies to slow the epidemic in the two transmission networks might be quite different. For example, suppose that spermicides such as nonoxonol-9 were found to be more effective in reducing β_i and α_i than in reducing β_r or α_r. Under this scenario, the use of spermicides could have a dramatic effect on the heterosexual spread but only a minor effect in the homosexual network, where, for example, condoms may be necessary to reduce both α_i and α_r.

Also, because other STDs may significantly raise the insertive infectivity β_i from a woman to a man, one of the most effective strategies for slowing the epidemic in the heterosexual network may be to launch a major campaign to reduce the incidence of other STDs. The recent dramatic increase (approximately 29% per year) of syphilis cases in the United States has been attributed by some to the transferring of STD educational and treatment dollars to fight the AIDS epidemic. This transfer may be a counterproductive approach and may result in a faster-spreading heterosexual epidemic. Once the relative infectivities are approximately known, then the model will be able to give guidance in answering questions such as whether it would be more effective to spend educational funds until, for example, 90% of the heterosexual contacts use condoms or to reduce the incidence of other STDs by 50% through contact tracing and treatment.

C. *DRUG USE*

HIV is transmitted by sharing needles to inject drugs. Partly because many prostitutes are drug users and partly because most drug users are

heterosexuals, the spread of HIV infection in the needle-sharing community is seen as a major source of HIV for the heterosexual community at large. Some important questions are what fraction of the population engages in needle sharing in different age groups and regions, how the drug users are distributed according to frequency of needle sharing, and how much bias exists toward sharing repeatedly with the same people and against sharing with strangers [19, 7]. All of the mixing questions raised about sexual activity also apply.

D. POPULATION DENSITY

The results from serological tests conducted by the Department of Defense (DoD) on potential recruits indicate that the prevalence of HIV is highly correlated with population density [9]. There are a number of reasons for this, each of which needs to be considered. Unlike many non-STDs (e.g., measles, influenza), the rate of infection should not be strongly dependent upon the density of the host; however, people in large cities are less constrained than those in small towns. Endemicity also plays a role, because the virus will be spread only when it is present. Finally, physical distance creates barriers between people, so mixing may be more random and homogeneous in denser areas. The spread of the virus into the regions surrounding the major population centers is a diffusionlike process in which the diffusion rate is a function of the population density.

E. ZONES

As mentioned above, isolation provided by distance provides another barrier to the epidemic. Behavior may also be somewhat regional. For example, the prevalence of shooting galleries in New York City may be a major reason why HIV has spread more rapidly within the New York City drug community than in the California drug communities, where shooting galleries are less common. In a risk-based drug-use model (as described in Section V.B), the partnership (needle-sharing) mixing distributions would be different for New York City than Los Angeles, and the predictions would be very different. Also, to understand how rapidly the HIV infection will spread into different regions of the country, we might want to model how each region is connected to every other region by the movement of people.

Infection through blood transfusions caused a wide geographic spread of the virus. In the spring of 1985, before stringent screening measures were applied to blood donors, 0.25% of the blood tested by the ELISA test was seropositive [10]. Infected blood led to a widespread scattering of HIV infections throughout the United States, which might have a major impact on the future course of the epidemic, even though only a tiny proportion of the population was infected this way. Today, most of the HIV-tainted blood

in the United States is identified by the ELISA test; therefore, current blood-transfusion infections may have a negligible effect on the course of the epidemic.

F. ETHNICITY AND SOCIAL GROUP

The number of AIDS cases that have occurred, especially those in women and children, are disproportionately greater in the Black and Hispanic populations than in the rest of the population [49]. The DoD data from military recruits also show this bias [9]. It is not understood why the infection has spread more rapidly into these populations. There are, however, social barriers to contacts between different racial groups, so it may largely be a question of endemicity. In other words, once the virus is introduced into a group of people, it can spread only in that group until a contact with a member of another group is encountered. If there are not enough contacts between racial groups, the virus can spread entirely in one group without extending into another. These groups need not be only racial; any isolated group with few outside contacts could experience an isolated spread. For example, students at the same university might form such a group. Lifestyle differences in these groups could result in different parameter values for the other risk factors.

We may divide individuals within a group into two classes: social and nonsocial. The nonsocial individuals interact only within their group, whereas the social individuals have contacts both within and outside their particular group. This approach has also been used to model other infectious diseases such as hepatitis [51, 52].

G. COFACTORS

Cofactors, such as diseases and practices that cause skin lesions or impairment of the immune system, may influence a person's susceptibility to becoming infected and, once infected, that person's infectiousness and/or disease progression. As yet the data on the effect of cofactors are poor, but numerous cofactors including syphilis, gonorrhea, herpes, drug use, and malnutrition have been proposed. These cofactors are more common in some groups, such as individuals in urban slums, than in others and could allow for more rapid spread in those groups than would occur in the absence of cofactors. For example, infectivity estimates from middle-class spouse/pair studies may not give correct estimates when other venereal diseases are present. It may be necessary to take account of the distribution of cofactors in the population to fully understand HIV spread.

In central Africa, cofactors probably account for the rapid heterosexual spread. Untreated genital ulcers, often caused by chanchroid, which are rare in developing nations, greatly increase infectivity.

III. DISEASE PROGRESSION

Studies of the long-term effect of the HIV virus on the immune system are all reaching similar conclusions: HIV causes a slow but progressive decline in the immune system. The rate of this decline varies from person to person, and some people appear to stay on a plateau for long periods. Short-term upward fluctuations in measurements of quantities such as the T-4 helper cells are often observed, but most infected immune systems decline over the long run [8, 48, 43]. Autopsies of AIDS victims show that HIV also crosses the blood brain barrier in a large percentage (around 80%) of infected persons and causes a wasting away of the brain; it is not yet clear if this deterioration is a slow progression or if it happens late in infection [16].

When the immune system is sufficiently compromised or when the brain is sufficiently affected, symptoms appear. Initial symptoms of immune problems range from the very mild (so-called AIDS-related complex [ARC], or generalized lymphadenopathy, or even just poor health) to KS and the devastating opportunistic infections classified as AIDS. Deterioration of the brain leads to blindness and Alzheimer's-like dementia. Eventually, death follows. It is not clear what the appearance of KS has to do with HIV-stimulated immune-system decline. KS may occur at any point more than 1 year after infection, independent of immune-system breakdown. It is much more prevalent in homosexual men from New York City than in other groups. It is often not the eventual cause of death; the immune-system decline continues until an opportunistic infection leads to death.

A. TIME FROM INFECTION TO AIDS

This picture of progressive immune-system decline indicates that most infected individuals eventually die from HIV-induced illness and that the probability that an individual will develop AIDS depends on how long he has been infected. Both the time from infection to diagnosis of AIDS and the time from diagnosis to death are extremely variable. HIV-infected adults have developed AIDS in less than 2 years, and some have remained well for more than 8 years. The distribution of times between infection and clinical AIDS is only partially known because of the long times involved. In studies of patients for whom an estimate of date of infection can be made (such as hemophiliacs), the percentages developing AIDS in any given year after infection are either still increasing or are remaining roughly constant, which leads to an estimate of an average time to AIDS of at least 8 years. A possible distribution of times from infection to clinical AIDS is shown in Figure 8 (Section VI.A).

Any model that is going to predict the number of infected people and AIDS cases must take into account the wide variability in duration of

infections proceeding AIDS. This is necessary to predict accurately the correct distribution of people developing AIDS and to ensure that infected people in the model remain infectious for lengths of times that reflect the actual infectious periods. A person who is healthy but infected for a long period has a higher probability of infecting someone else than a person who develops AIDS relatively early.

Similarly, there is a wide distribution of times between diagnosis of AIDS and death. Death may occur immediately after diagnosis or more than 6 years later, with an average patient lifetime of 12–14 months. In the future, these times will depend strongly upon the effectiveness of therapy such as the use of AZT. Keeping track of time since infection allows us to use "best guess" estimates for these distributions. Another effect of the long duration of infection is that as infection progresses, people will ascertain their seropositivity and change their behavior.

B. VARIABLE INFECTIVITY

Infectiousness of individuals carrying HIV varies as the course of the disease progresses. In studies of infected hemophiliacs and blood transfusion recipients, few of their spouses have seroconverted sooner than a year before the infected individuals developed AIDS or ARC symptoms [21, 17]. This time lag indicates that infectiousness is often minimal until late in the course of the infection. However, some partners have been known to convert immediately [56].

The infectivity may be related to the amount of free virus in the circulatory system of an infected individual. Studies indicate that the amount of free virus goes up in the first few weeks after infection [18, 54] and then goes down as antibody response occurs, remaining at very low levels for years. There may be sporadic bursts of free virus and hence of infectivity in these intermediate years because of other challenges of the immune system. As the immune system collapses in the year or so before AIDS develops, viral counts return to high levels (Robert Redfield, private communication; [36]). This progression is schematically shown in Figure 2.

It is clear both from the infectivity studies of Grant et al. [24] and Padian et al. [45] and from our numerical studies that the chances of infection from a single sexual contact must be quite low (less than about 0.01) for most of the duration of infection, or else the virus would have spread much faster than it has. If the initial infectious period does exist; it is important that it be well defined for infected individuals with many contacts, because it has a large effect on the rapid-growth phases of the epidemic. This effect is especially strong when a disproportionately large percentage of infected people have only just become infected.

Such a radical time variability of infectiousness raises an additional possibility. We know that the number of infected people has grown rapidly

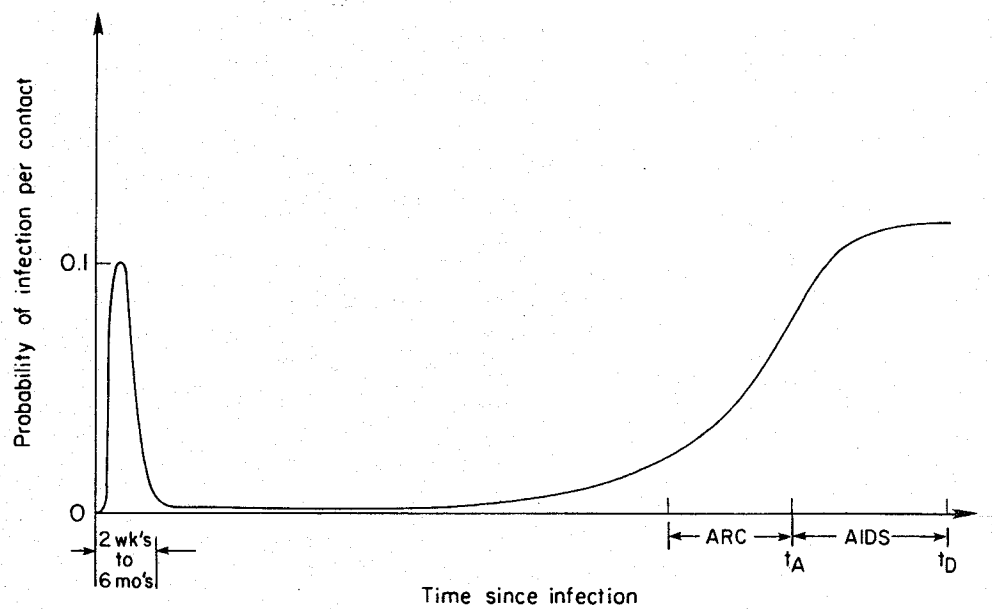


FIG. 2. Schematic of the infectiousness of a sexual contact with the same individual. Initially, the virus quickly multiplies but then is suppressed by the immune system. Towards the end of infection, viral counts again become high, coincident with immune-system breakdown. This infectiousness curve is based on relative amounts of HIV in the infected individual. Also, spouse/pair studies indicate the infectiousness of an individual must be low during the first few years after the initial immune response.

during the early stages of the epidemic, with doubling times significantly less than a year. It may be that infection has primarily been transmitted from the infected to the noninfected in the early time interval of roughly 2–6 months, with the period of low infectiousness, 0.7–5 years in Figure 9 (Section VI), contributing to fewer total infections. The periods of increasing infectivity, 1–3 years before AIDS, have a reduced relative contribution because of the rapid growth of the infected population during the low-infectious period. This reduced contribution may be especially true because the people with large numbers of partners were infected first and could have encountered many people during the initial period. However, as the growth of the epidemic slows, and the epidemic moves into groups with less than 1 partner per 6 months, contacts with people in the later disease stages will become the primary transmission route.

Because the disease is much more infectious in the later stages, widespread screening and voluntary testing to identify the HIV carriers (before they enter this stage) will be more effective than if the infectivity were constant. Any cost/benefit analysis for testing must take variable infectivity into account. Changing an individual's behavior before he or she enters this very infectious stage could be one of the most effective means of slowing the epidemic.

C. CLINICAL MANIFESTATIONS

Models could differentiate between the various clinical manifestations of AIDS based on different conversion probabilities. At this time we do not differentiate but have a lumped conversion-probability distribution, which peaks between 7 and 10 years and assumes every infected individual eventually converts to AIDS. The conversion time may be longer in healthier and younger populations, and medical advances may lengthen the conversion time.

D. GENETIC VARIATION

The genetic variability of HIV DNA sequences indicates that the virus is mutating 5 to 10 times faster than an influenza virus [53, 25]. The variability is due primarily to duplications, insertions, or deletions of short segments and point mutations. The various strains may have dramatically different resistance to vaccines or may lead to different etiologies (e.g., dementia, KS versus PCP). If differential viral strains have different etiology, then some strains may eventually win out over others. For example, strains with longer incubation times, those that are more infectious, strains such as HIV-2 that are not recognized by the ELISA test, or those that least reduce the health of the infected person when they are most infectious may eventually spread faster than other strains.

IV. ANALYTIC FORECASTING

The accumulated number of AIDS cases diagnosed in the United States as reported to CDC, $A(t)$, is not growing exponentially but is well approximated by the cubic polynomial

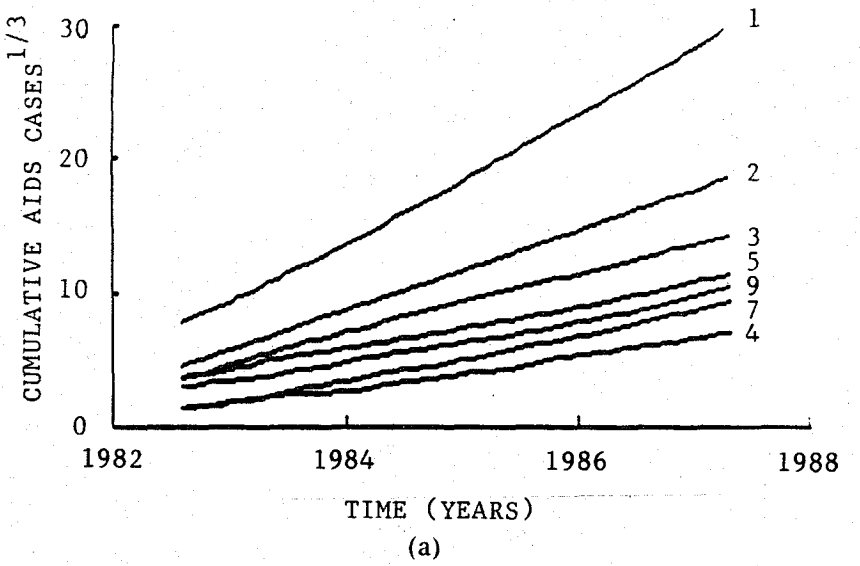
$$A(t) = 174.6(t - 1981.2)^{3.0} + 340 \pm 2\% \tag{4.1a}$$

for times $t \geqslant 1982.5$. The rate of new AIDS cases per year, $A'(t)$, is similarly approximated by the derivative of Equation (4.1a), the quadratic equation

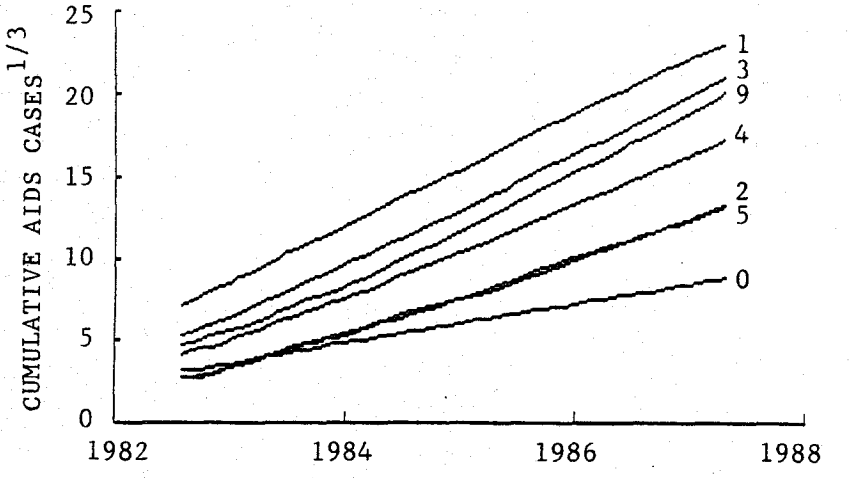
$$A'(t) \simeq 523.8(t - 1981.2)^{2.0}. \tag{4.1b}$$

This polynomial growth is evident in nearly every CDC-defined category, including risk behavior [Figure 3(a)], age, region of the country [Figure 3(b)], and ethnic group [30]. The AIDS cases approximated by Eq. (4.1) are based on the pre-June 1987 AIDS definition and do not include dementia and wasting syndrome.

Because the cumulative growth of AIDS cases is cubic [Equation (4.1a)], the cube-root reference frame shown in Figure 3 is a natural frame to identify changes in the epidemic. Similarly, the incidence data should be



(a)



(b)

FIG. 3. (a) The $\frac{1}{3}$ power of the CDC AIDS cases is linear in risk categories 1–5, indicating an approximate cubic growth, $a(t-t_0)^3$. The risk categories are 1, homosexual male; 2, IV drug user; 3, homosexual and IV user, male; 4, hemophiliac; 5, heterosexual contact; 7, transfusion; 9, other or unknown. Here the lags in reporting time have been approximated and corrected. (Data analysis by C. Qualls.) (b) The $\frac{1}{3}$ power of the CDC AIDS cases is linear in different regions of the United States, indicating an approximate cubic growth, $a(t-t_0)^3$. The metropolitan regions are 1, northeast; 2, central; 3, west; 4, south; 5, mid-atlantic. Region 9 is nonmetropolitan, and region 0 is unknown. Note that the nonmetropolitan AIDS cases in region 9 did not grow as a cubic until 1984. (Data analysis by C. Qualls.)

studied in the square-root reference frame. The data are also linear in a log-log reference frame where time has been shifted so $t = 0$ corresponds to 1981.2. In the log-log reference frame the exponents can easily be determined by the slope of a linear least-squares fit. If the data are plotted in a log-linear reference frame, then extrapolation of future cases becomes much harder and anomalies such as for region 9 in Figure 3(b) are less evident.

Because the growth is polynomial [Equation (4.1a)], the doubling time is not constant but is increasing linearly; setting $A(t + t_d) = 2A(t)$ in Eq. (4.1a) defines the doubling time $t_d \simeq 0.26(t - 1981.2)$ years. This increasing doubling time has led some observers analyzing the data in a log-linear reference frame to incorrectly state that the epidemic is leveling out, when in the cube-root reference frame (Figure 3) it is clear that the trends have been consistent for the past 5 years.

The cubic polynomial growth can be explained by a wave of infection progressing from populations with high-risk behavior into populations with lower-risk behavior. For example, if individuals with risk behavior r (proportional to the number of sexual partners or needles shared) are infected through interactions with people of similar behavior and if the population is distributed as a decreasing function of risk behavior [e.g., $N(r) \simeq N_0(1 + ar)^{-4}$, where $N(r)$ is the number of individuals with risk r], then the highest-risk population is quickly infected, giving rise to an initial transient exponential growth. This growth quickly becomes polynomial as the saturation wave of infection moves into lower-risk (but still high-risk) behavior and finally slows to an $e^{1/t}$ growth rate (see Section VII.B). The polynomial growth is analyzed in more detail in [11].

If $C(\tau)$ is the probability that a person infected with HIV at time $t - \tau$ has developed AIDS by time t, and if $I'(t)$ is the number of people infected per year with HIV, then the cumulative AIDS cases reported to CDC satisfies the relationship

$$A(t) = p\int_0^\infty C(\tau)\, I'(t - \tau)\, d\tau, \tag{4.2a}$$

or, because $C(0) = 0$,

$$A'(t) = p\int_0^\infty C'(\tau)\, I'(t - \tau)\, d\tau, \tag{4.2b}$$

where p is the fraction of infected individuals eventually reported to CDC as AIDS cases. Thus p is the product of the probability that an infection will result in a pre-1987.5 CDC-defined AIDS case (which excludes dementia and slim disease) times the probability it will be reported to CDC. The probability that an AIDS case will be reported to CDC is the product of the probabilities that it will be diagnosed and, once diagnosed, that it will then be reported. Using estimates of $C'(\tau)$, the probability density function for

conversion to AIDS τ years after infection, we can solve Equation (4.2) for $I'(t)$. In these calculations we used $p = 0.72$ and the Weibull probability density function for $C'(\tau)$ described in Section VI [Figure 8, Equation (6.1)], unless otherwise stated.

Solving equation (4.2) for $I'(t)$ is ill posed; small changes in $A(t)$ or $C(\tau)$ may cause large changes in $I'(t)$. We solved both Equations (4.2a) and (4.2b) by a least-squares quadrature method where $I'(t)$ was approximated by piecewise cubic Hermite polynomials (splines). $A(t)$ was extrapolated using Equation (4.1). The calculated solutions agreed within 5% when 10 to 30 piecewise polynomials were used. Below 10 piecewise polynomials the approximation was too coarse, and above 30 the ill-posed nature of the problem created high-frequency oscillations in the solution.

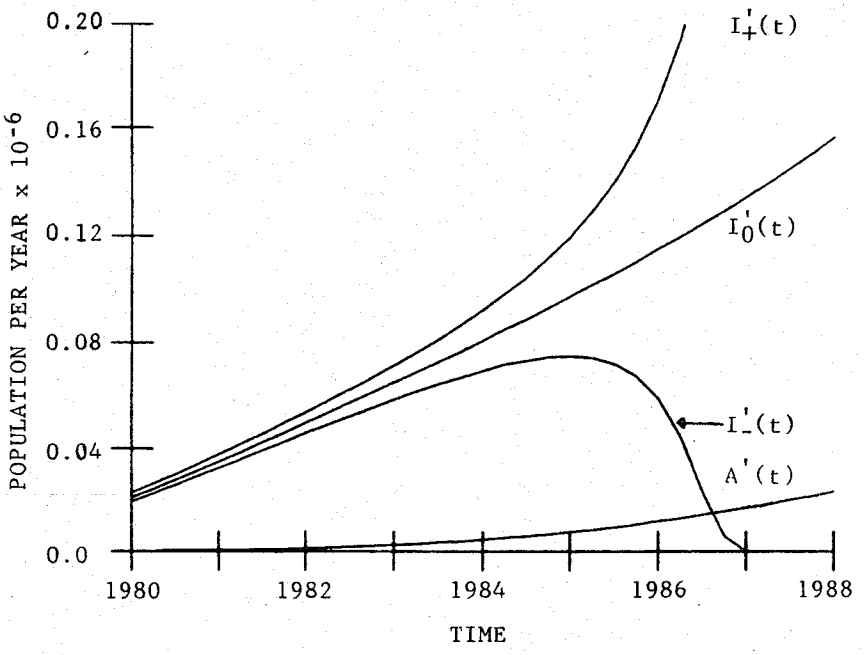
The cumulative number of infected individuals, $I(t)$, was most sensitive to the extrapolated estimates of $A(t)$, the fraction of the infected population that eventually is reported to CDC as AIDS cases, and the most likely conversion time to AIDS, which we call τ_A $[C''(\tau_A) = 0]$. After the initial transients, $I(t)$ was relatively insensitive to the width of the distribution $C'(\tau)$ about τ_A. The uncertainty that an HIV infection will result in an AIDS case reported to CDC is a linear factor and changes the estimates for the infected by p^{-1}. We now investigate the effects that these uncertainties in each of $C(\tau)$, the fit of Equation (4.1), and the extrapolation of Equation (4.1) have on the estimates for the infected population.

Because so few people develop AIDS in the first 2 years after infection, it is clear that today's AIDS cases cannot be used to estimate the number of people infected in the past 2 years with any accuracy at all. In fact, the error estimates for I' near $(t_0 - \tau)$ explode proportional to the relative error in $A(t_0)$ times $C^{-1}(\tau)$. Here t_0 is the maximum time at which $A(t)$ is specified. More generally, a relative error in $I'(t_0 - \tau)$ of

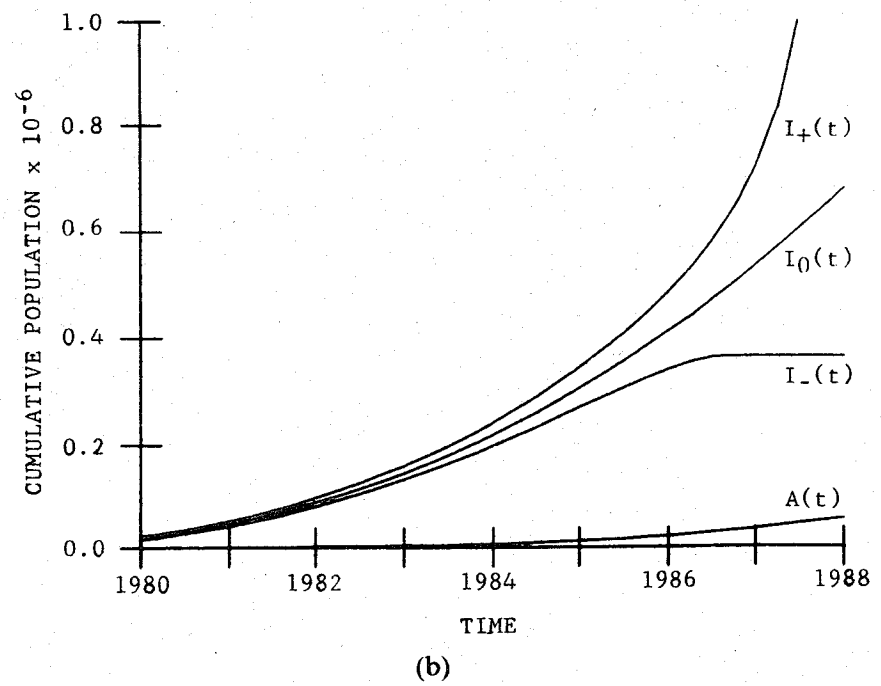
$$\epsilon_I(\tau) = \epsilon_A C^{-1}(\tau)[1 - C(\tau)]\left[I(t_0)A^{-1}(t_0) - 1\right]^{-1}$$

can be introduced in Equation (4.2) and the relative values of $A(t_0 - \tau)$ will change less than ϵ_A for $\tau \geq 0$. In Figure 4 we plot the solution $I_0(t_0 - \tau)$ of (4.2) using (4.1). Also plotted are the upper and lower error bounds, $I'_{\pm}(t_0 - \tau) = I'_0(t_0 - \tau)[1 \pm \epsilon_I(\tau)]$, for $\epsilon_A = 0.01$, $t_0 = 1988$. These error bounds for $I'(t)$ and hence $I(t)$ due to errors in $A(t)$ are small for $t \leqslant 1984$ but gradually increase and explode between 1 and 3 years ago.

The upper and lower bounds I'_+ and I'_- for $I'(t)$, $t \leqslant t_0$, in Figure 4 would result in very different future values for $A(t)$, $t \geqslant t_0$. To reduce the error bounds on $I'(t)$, $t \leqslant t_0$, we must incorporate assumptions on the behavior of $A(t)$, $t \geqslant t_0$. That is, to estimate the number of infecteds at time t_0, we must first estimate future AIDS cases for times up to, say, $t = t_0 + (1/2)\tau_A$. These extrapolations can then be used in Eq. (4.2) to estimate the infected population $I(t)$ for $t \leqslant t_0$.



(a)



(b)

FIG. 4. (a) The upper and lower error bounds I'_+ and I'_- for $I'(t)$ from the inversion of Equation (4.2) using the Weibull conversion-time distribution of Medley et al. [42] explode as t approaches 1987 when $A(t)$ is not extrapolated beyond 1988.0. These three estimates for I' predict cumulative AIDS cases that agree within 1% for $t \leqslant 1988$ but are very different for $t > 1988$. (b) The error bounds for the cumulative number of infecteds, $I(t)$, corresponding to the integrals of I' in (a).

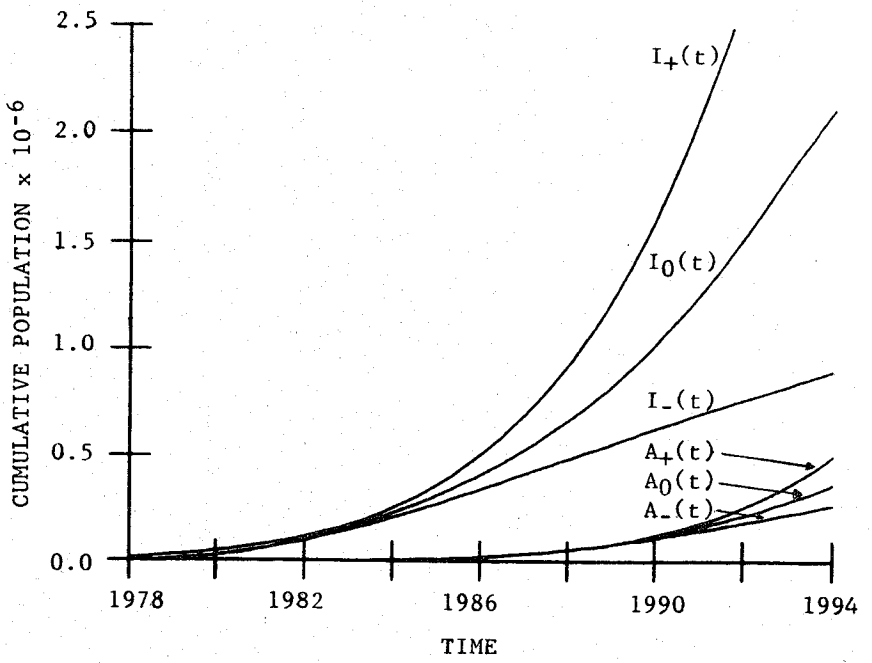


FIG. 5. The rate of new AIDS cases was extrapolated using (4.1b) with a maximum relative error of $\pm 10\%$ per year after 1988. The cumulative infected population $I(t)$ was then estimated for $t \leqslant 1994$. Here $p = 0.72$ and $C'(\tau)$ is defined in Figure 8 (Section VI).

Even though the cubic extrapolation in Equation (4.1) closely agrees with the data over the past 5.5 years, it is purely an empirical fit to these data. This approximation is not based on transmission mechanisms and therefore does not include any effects of behavior changes that are known to have occurred, saturation of infections in certain risk groups, the screening of the blood supply, the infection starting in new populations, or any other of the major influences on the future course of the epidemic. Because the underlying transmission dynamics are changing, we do not expect the cubic to continue to hold indefinitely, and hence the traditional statistical confidence bounds are not an appropriate tool to estimate errors in forecasting future AIDS cases. Therefore, we have kept the error analysis simple in our investigations of the sensitivity of the estimates of the infected population on the future-AIDS-case projections. To allow for a relative error of ϵ per year in the rate of AIDS cases per year, we multiply Equation (4.1b) by $(1 \pm \epsilon)^{t-1988}$ for $t \geqslant 1988$. In Figure 5 we demonstrate the sensitivity of the estimated infected population to future AIDS predictions by solving (4.2) for $I_{\pm}$ with $A'_{\pm}(t) = A'(t)(1 \pm 0.01)^{t-1988}$.

To investigate the effects of the uncertainties in $C'(\tau)$, we extrapolated the AIDS cases using Equation (4.1) and compared the solutions of Equation (4.2) in Figure 6 for four different conversion functions: (a) the Weibull distribution shown in Figure 8 (Section VI), which has a most likely time of

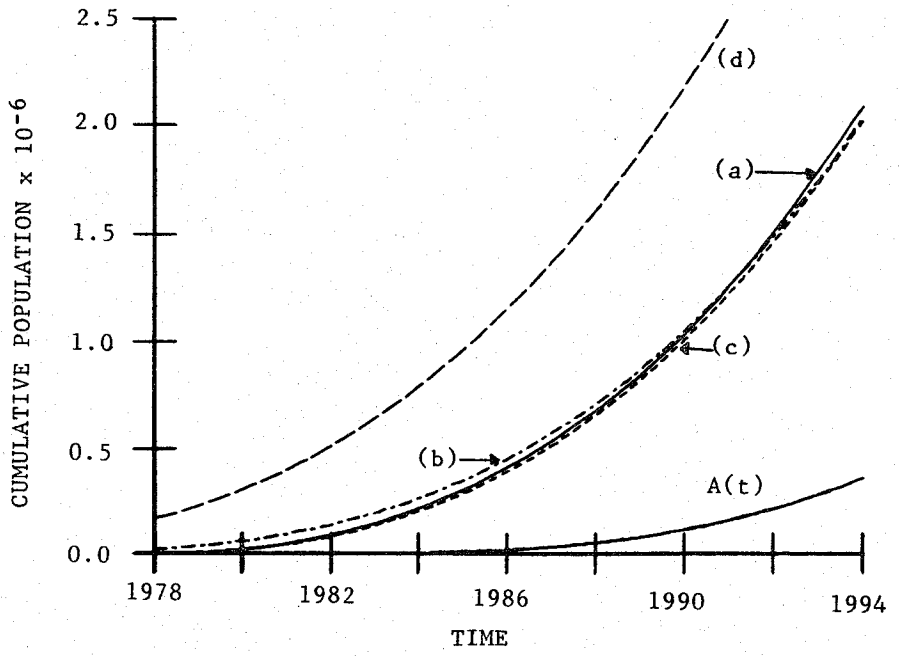


FIG. 6. Equations (4.2) were solved by extrapolating the cumulative AIDS cases in Equation (4.1) and setting the reporting fraction $p = 0.72$. After some initial transients, the solutions are insensitive to the shape of the conversion-time distributions as long as the most likely conversion times agree. (a) (solid line) Weibull; (b) (short dashes) δ-function at 7.5 years; (c) (dash-dot) step function, 2–14 years; (d) (long dashes) δ-function at 12 years.

$\tau_A = 7.5$ years and a median time $[C(\tau) = \frac{1}{2}]$ of 8 years; (b) a delta function at 7.5 years (that is, everyone infected develops AIDS in exactly 7.5 years); (c) a step function that is 0.083 for τ between 2 and 14 years and 0 otherwise (there is no most likely time; the median time is 8 years); and (d) a δ-function at 12 years. The solutions for the number of infecteds for the first three of these agree within a few percent, as shown in Figure 6.

As the width of $C'(\tau)$ approaches zero (that is, a δ-function), then the solution of Eq. (4.2) approaches.

$$I(t) = p^{-1}A(t + \tau_A). \tag{4.3}$$

This estimate can be used as a rough approximation for $I(t)$, even for fairly wide distributions $C'(\tau)$, as demonstrated in Figure 6. This approximation can be used to estimate the number of infected individuals in January 1988. For example, if we assume that 80% of the infected individuals develop CDC-defined AIDS and that 90% of these are reported to the CDC, then $p = 0.72$. If $\tau_A = 9$ years and the number of AIDS cases in 1997 ($= 1988 + \tau_A$) is 85% of the extrapolated cubic approximation (4.2), then the current

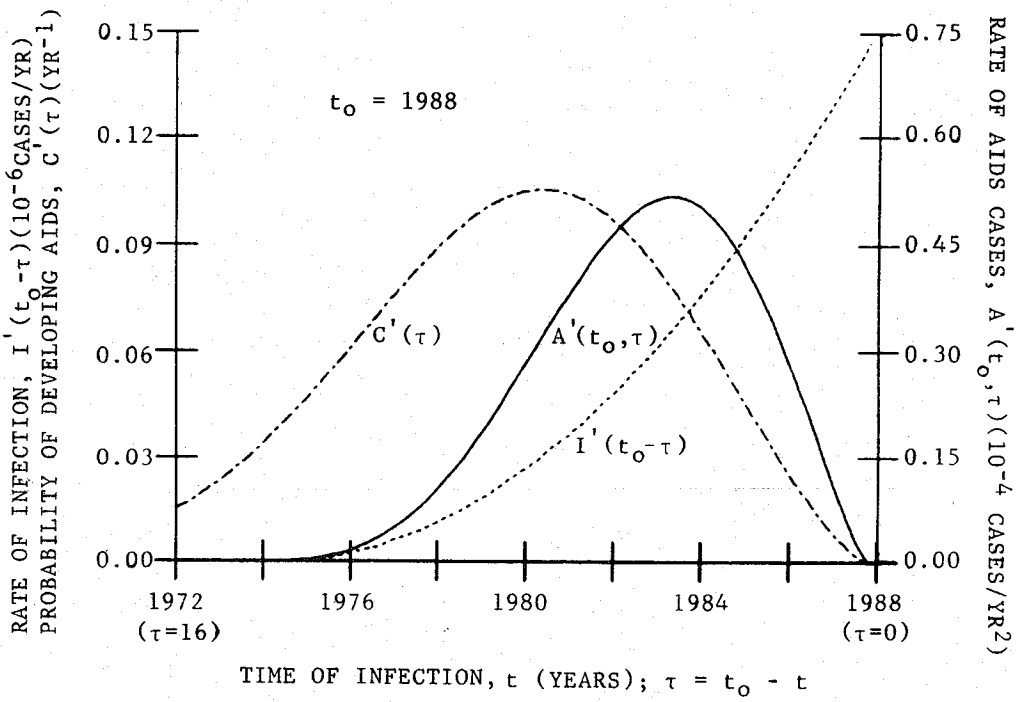


FIG. 7. Because the number of people infected with HIV is growing rapidly, the distribution of time since infection in a group diagnosed with AIDS in any time period does not reflect the distribution of times from infection to AIDS. Multiplying $I'(t_0-\tau)$ by $C'(\tau)$ gives the distribution of AIDS cases, $A'(t_0,\tau)$, occurring at time t_0 that were infected at time $t_0-\tau$. Note that because $I'(t)$ is increasing, the most likely time to AIDS ($7\frac{1}{2}$ years) is longer than the most likely length of infection for current AIDS cases ($4\frac{1}{2}$ years).

cumulated number of infected individuals is

$$I(1988) \simeq \frac{0.85}{0.72}\left[174.6(1988.0+9-1981.2)^3+340\right] \cong 810{,}000. \quad (4.4)$$

We remark that if only 40% of the infected individuals develop CDC-defined AIDS (as was thought a few years ago) then, even though the predicted AIDS cases are the same, this approximation estimates that there would be 1,600,000 people infected with HIV in the United States.

Although the numbers of AIDS cases for different conversion-time distributions in Figure 6 agree, the length of times that the current AIDS cases have been infected would have very different distributions. Most of the AIDS cases diagnosed today were infected 3–5 years ago. This time is shorter than the most likely time to convert to AIDS because the HIV-infected population is growing rapidly. In Figure 7 we give an example where the rate of growth for the HIV-infected population $I'(t)$ is quadratic and the most likely time to convert to AIDS is 7.5 years. The distribution of patients currently developing AIDS, $A'(t_0,\tau)$, is the product of the two dashed curves, $C'(\tau)I'(t_0-\tau)$, and hence is highly skewed toward the early conversion times.

The current most likely time since infection for current AIDS cases in Figure 7 is $4\frac{1}{2}$ years. This would imply that we should soon see a slowing of AIDS cases in transfusion recipients. Also, the sexual-behavior changes in 1983–84 in the San Francisco area should just now be reducing the growth in the homosexual cases.

Note that, even when the time since infection is known, $A'(t, \tau)$ is insufficient to determine $C'(\tau)$ unless an estimate of $I'(\tau)$ can be made. For the transfusion-infected patients, the shape of $I'(\tau)$ could be ascertained if the contaminated fraction of the blood supply were known as a function of time.

Simplified analytic forecasting models such as these can give good estimates of how many people will get AIDS next week and give fair estimates for the next year, but they are insufficient to accurately predict the course of the epidemic 3–5 years from now. The only way to make reliable long-term predictions is to include far more detail on the epidemiology and sexual behavior through full-scale computer models.

V. SIMPLIFIED MATHEMATICAL MODELS

A complete model of the spread of the AIDS virus in a sexually active and IV-drug-using community must account for the complicated interactions between people. However, one must begin by understanding the behavior of simple models before going on to explore more complex ones. Two different approaches to this modeling have been developed. One is based on ascertaining the risk to an individual [28, 3]. The other is based on population-growth models in which individuals form and break partnerships. In this approach, paired individuals become infected through multiple contacts when one partner is infected, but remain protected for the duration of the partnership if both are uninfected and also cannot become infected between partnerships [14, 13]. In the risk-based models, the population is easily stratified according to the amount of risk individuals incur, but they do not model well the risk (or protection) of longer-term relations. On the other hand, the partnership models are more difficult to stratify so as to take account of the wide variations in risk behavior within the population. In this paper, we are primarily concerned with modeling HIV spread in high-risk populations, so we use the risk-based approach. We do, however, account partially for partnership duration by allowing a variable number of contacts in each partnership.

In this section, we describe several models for the sexual spread of HIV in a population structured only according to the rate of acquisition of new partners; similar models would hold for needle-sharing associated with IV drug use. We begin with a model that neglects all heterogeneities in the susceptible and infected populations, present a similar model for heterosex-

ual spread, address the variations due to progression of infection, and finally split the population according to risk behavior. Parameter estimates for these models will be discussed in Section VI, and their behavior will be explored in Section VII.

The models discussed here do not incorporate behavior changes, although that is easily added. We assume that this population would be stable if HIV were not present, with migration and maturation into the group balancing deaths and aging processes that remove people from the group. We assume there is no immunity, so all uninfected members are susceptible to infection. Once infected, a person remains infected, infectious, and sexually active until AIDS intervenes. In all but a simple heterosexual model we do not distinguish the sex of the members of this at-risk community. The assumptions involved in the development of these models will be described in more detail in a later report where a derivation of the probability-of-infection function is given.

A. τ-MODELS

In our models, we divide this at-risk community into uninfected susceptibles, infecteds without AIDS, and diagnosed AIDS cases. We assume that before the introduction of the AIDS virus, there was a balance between a constant maturation and migration rate into the community and a constant rate per individual of retirement or death out of it; these processes continue in the presence of AIDS. Susceptibles become infected through sexual contacts or IV needle-sharing with partners whom they choose randomly, at a fixed rate, from the susceptible and infected portions of the community. Infected individuals eventually develop AIDS, become sexually (or needle-sharing) inactive, and die at an accelerated rate.

In the simplest model where

t = time,
$S(t)$ = number of susceptible individuals,
$I(t)$ = number of infected individuals without AIDS,
$A(t)$ = number of AIDS cases,
$A_T(t)$ = accumulated number of AIDS cases,
$N(t)$ = number of susceptible and infected individuals without AIDS,
μ = death rate of individuals without AIDS,
δ = death rate of individuals with AIDS,
γ = rate of developing AIDS of infected individuals,
i = probability of infection from a sexual contact with an infected,
c = average number of contacts between sexual partners,
r = average number of new sexual partners per year, and
S_0 = population size before the AIDS virus was introduced.

we have

$$\begin{aligned}
\frac{dS(t)}{dt} &= \mu(S_0 - S(t)) - \lambda(t)S(t), \\
\frac{dI(t)}{dt} &= \lambda(t)S(t) - (\gamma + \mu)I(t), \\
\frac{dA(t)}{dt} &= \gamma I(t) - \delta A(t), \\
\frac{dA_T(t)}{dt} &= \gamma I(t),
\end{aligned} \tag{5.1}$$

with

$$\lambda(t) = icr\frac{I(t)}{N(t)},$$

$$N(t) = S(t) + I(t),$$

where $\lambda(t)$ is the rate of infection per susceptible. [Note that we changed notation from Section IV, where $I(t)$ and $A(t)$ were cumulative quantities, rather than current numbers as they are here.]

This model portrays a community in which people mature or migrate into the susceptible community at a constant rate μS_0. People without AIDS die (or become inactive) at a constant rate, with μ^{-1} their average life expectancy. Infection occurs through sexual contact with an infected partner. Partners are chosen at random, from all susceptibles and infecteds, at an average rate r per year, so that the probability a partner is infected is the fraction of infecteds in the population. Infecteds develop AIDS at a rate γ, and AIDS cases have a decreased life expectancy δ^{-1}.

This simple model has been presented and analyzed by many authors for the spread of various sexually transmitted diseases, including AIDS (see, for example, [28, 3]).

Equation (5.1) can be modified to model purely heterosexual spreading by splitting the population according to sex and including the partnership balance relationships. These balances are necessary to take account of situations where there are not enough women so that men cannot have as many partners as they might like, and vice versa. A symmetric model that satisfies the condition that the number of female partners equals the number

partners, and where we define

$S_0^T = (w_0, m_0)$: initial populations of women and men,
$S^T = (w(t), m(t))$: susceptibles,
$I^T = (W(t), M(t))$: infecteds,
$A^T = (W_A(t), M_A(t))$: AIDS cases,
$r^T = (r_w, r_m)$: average desired number of male (r_m) or female (r_w) partners the women or men have per year,
$I^T = (i_w, i_m)$: infectiousness of a contact with a woman (i_w) or a man (i_m),

again gives the system in Equation (5.1). Here we have used the superscript T to indicate the transpose operator. The rate of infection per susceptible is

$$\lambda(t) = (i_m M, i_w W)^T c r_m r_w \left[(S+I)^T r^T\right]^{-1}, \tag{5.2}$$

and the product λS in Equation (5.1) is defined as $(\lambda_1 w, \lambda_2 m)^T$.

The initial growth of infecteds in both of these models is exponential, with

$$I(t) = I_0 e^{(\alpha-\gamma-\mu)t},$$

where $\alpha = icr$ for the model in Equation (5.1) and $\alpha = cr_w r_m (i_w i_m w_0 m_0)^{1/2} (r_1^T S_0^T)^{-1}$ for the model in Equation (5.2).

The simplistic choice of λ in Equation (5.2) illustrates the complexity introduced by balancing male and female partners. Adding additional structure, such as risk behavior or age, complicates the balance equations even more. Currently, there is very little information on male-female mixing patterns. If the solutions are sensitive to the different balance equation assumptions, then more data will be essential.

These two models assume, among other things, that all contacts with infected persons are equally infectious throughout the course of infection. As discussed in Sections III.B and VI.B, there may be a wide variation in infectiousness as the disease progresses. The constant rate of progressing to AIDS imposes an exponentially decaying distribution of times to AIDS. However, cohort studies have found that the probability of getting AIDS increases with time since infection for at least the first 7 years (see Section VI.A). Thus, a decaying distribution is a poor approximation.

If we include time since infection or AIDS, then variable infectivity and the distributions of times from infection to AIDS and of times from AIDS to death may be explicitly modeled. Following Anderson et al. [3], we break down the infected population $I(t)$ according to the time τ since infection, $I(t) = \int I(t,\tau)\,d\tau$. $I(t,0)$ is now the rate at which people become infected, and $I(t,\tau)$ has the units people/year. Similarly, we distribute AIDS patients

according to time τ since AIDS, $A(t) = \int A(t,\tau)d\tau$. Defining

$I(t,\tau)$ = distribution of infecteds according to time τ since infection. $I(t,\tau)$ is the number of people infected per year τ years before time t,

$A(t,\tau)$ = distribution of AIDS cases according to time since they developed AIDS,

$i(\tau)$ = probability of infection from a contact with a person infected τ years ago.

$\gamma(\tau)$ = rate of developing AIDS at a time τ after infection,

$\delta(\tau)$ = death rate at time τ after developing AIDS,

we have the system

$$\begin{aligned}
\frac{dS(t)}{dt} &= \mu(S_0 - S(t)) - \lambda(t)S(t), \\
I(t,0) &= \lambda(t)S(t), \\
\frac{\partial I(t,\tau)}{\partial t} + \frac{\partial I(t,\tau)}{\partial \tau} &= -[\gamma(\tau) + \mu]I(t,\tau), \\
A(t,0) &= \int_0^\infty \gamma(\tau)I(t,\tau)\,d\tau, \\
\frac{\partial A(t,\tau)}{\partial t} + \frac{\partial A(t,\tau)}{\partial \tau} &= -\delta(\tau)A(t,\tau), \\
\lambda(t) &= \frac{cr}{N(t)}\int_0^\infty I(t,\tau)i(\tau)\,d\tau, \\
N(t) &= S(t) + \int_0^\infty I(t,\tau)\,d\tau \\
\frac{dA_T(t)}{dt} &= A(t,0).
\end{aligned} \tag{5.3}$$

The infectivity, $i(\tau)$, is an average over all individuals infected at time τ and is discussed in more detail in Section VI.B. For constant γ, i, and δ, $I(t)$ and $A(t)$ satisfy Equation (5.1). Although we have not done so, it would be easy at this point to vary c and r with time since infection and to thus take account of behavior changes caused by infection. For transmission in a heterosexual population, the model in Equation (5.2) is generalized in the same manner, with the infecteds, I^T, distributed according to time since infection and the AIDS cases distributed by time since diagnosis.

B. RISK-BASED MODELS

So far, the models presented do not treat variations in risk behavior between different people in the group. These models would be sufficient if the variation in risk behaviors were not large and did not play such a significant role in the epidemic. However, surveys of risk behaviors in the homosexual communities demonstrate that the variance in the number of

sexual partners per year is large. For example, the data for London in 1985–86 have a mean of around 25 partners/year and a variance of roughly 75 (see Section VI.D).

In this epidemic, it is significant that the people with many partners tend to become infected first and then become carriers who infect less-active people. This distribution can have a marked effect on the course of the epidemic and on which risk group is currently at highest risk of infection.

To model risk behavior, we suppose that the population can be distributed according to their numbers of new sexual partners per year. People mature into a fixed risk group and leave it only at death. Letting

r = number of new sexual partners per year,
$S(t,r)$ = distribution of susceptibles according to the number of partners per year,
$I(t,r,\tau)$ = distribution of infecteds according to the number of partners per year and the time since infection,
$c(r,r')$ = total number of contacts in a partnership between people with r and r' partners per year, and
$S_0(r)$ = density of people with r new partners per year before the AIDS virus was introduced,

we have the model

$$\begin{aligned}
\frac{\partial S(t,r)}{\partial t} &= \mu[S_0(r) - S(t,r)] - \lambda(t,r)S(t,r),\\
I(t,0,r) &= \lambda(t,r)S(t,r),\\
\frac{\partial I(t,\tau,r)}{\partial t} + \frac{\partial I(t,\tau,r)}{\partial \tau} &= -[\gamma(\tau)+\mu]I(t,\tau,r),\\
A(t,0) &= \int_0^\infty \int_0^\infty \gamma(\tau) I(t,\tau,r)\, d\tau\, dr,\\
\frac{\partial A(t,\tau)}{\partial t} + \frac{\partial A(t,\tau)}{\partial \tau} &= -\delta(\tau)A(t),\\
\frac{dA_T}{dt} &= \int_0^\infty \int_0^\infty \gamma(\tau) I(t,\tau,r)\, d\tau\, dr,\\
\langle rN(t)\rangle &= \int_0^\infty rN(t,r)\, dr\\
N(t,r) &= S(t,r) + \int_0^\infty I(t,\tau,r)\, d\tau.
\end{aligned} \tag{5.4}$$

We must still define $\lambda(t,r)$. We discuss below two possible choices: random partner choice and a bias of people towards partners like themselves. Note that now $S(t,r)$ and $S_0(r)$ have the units people·time/partners, and $I(t,\tau,r)$ has the units people/partners.

Random Choice. If we assume that partners are chosen at random from the entire population, then $\lambda(t,r)$ is given by

$$\lambda(t,r) = \frac{r}{\langle rN(t)\rangle}\int_0^\infty c(r,r')\,r'\int_0^\infty i(\tau)\,I(t,\tau,r')\,d\tau\,dr'. \qquad (5.5)$$

This model, except with no differences in partnership durations and no variability in infectiousness [$c(r,r')$ and $i(\tau)$ constant], was first proposed by Anderson et al. [3]. It assumes that the average $r-r'$ partnership is sufficiently short and infectivity is sufficiently low that the probability that a person has already become infected in the partnership is small, i.e.,

$$\max_\tau \{\, i(\tau)\}\, c(r,r') \ll 1.$$

Furthermore, the epidemic cannot grow so fast that the chance that a partner is infected becomes significantly different during the course of the partnership from an unmatched person from the same risk group. Anderson et al. [3] show that the initial growth of this model is determined not by the average number of partners/year, $\bar{r}$, but instead by $\bar{r}+\sigma^2/\bar{r}$, where σ^2 is the variance about this mean. They then proceed to approximate the model in Equation (5.5) by replacing r with $\bar{r}+\sigma^2/\bar{r}$.

Biased Partner Selection. The $\lambda(t,r)$ given by Equation (5.5) takes no account of the fact that people do not choose partners at random from all groups but instead prefer partners of a certain type and choose them when available. Ideally, the partner selection in any model should be based on sociological data. This question will be discussed in more detail in a later report; as a first step towards addressing this question we present below a model with a strong bias of people toward partners of similar risk behavior.

If mixing occurred only with people from the same risk group, then the virus could not spread between groups and the system in Equation (5.1) would describe separate epidemics for each value of r. However, this perfect isolation is unrealistic. The mixing between people of similar, but not identical, risk behavior leads to diffusion of the virus from one group to another. Under the assumptions described below, the rate of infection of a susceptible with risk behavior r, $\lambda(t,r)$, is approximated by

$$\lambda(t,r) = \phi(t,r)\left[1-\frac{\epsilon r}{N(t,r)}\left(5\frac{\partial N(t,r)}{\partial r}+r\frac{\partial^2 N(t,r)}{\partial r^2}\right)\right]$$
$$+\epsilon r^4\left[5\frac{\partial}{\partial r}\left(\frac{\phi(t,r)}{r^2}\right)+r\frac{\partial^2}{\partial r^2}\left(\frac{\phi(t,r)}{r^2}\right)\right], \qquad (5.6)$$

where

$$\phi(t,r) = \frac{c(r,r)}{N(t,r)} \int_0^\infty i(\tau)\, I(t,\tau,r)\, d\tau.$$

This expression for $\lambda(t,r)$ is derived as follows. Under the same assumptions on partnership duration, infectivity, and epidemic growth rates as mentioned above, $\lambda(t,r)$ can be approximated by

$$\begin{aligned} \lambda(t,r) &= \int_0^\infty p(t,r,r')\, k(t,r,r')\, dr' \\ k(t,r,r') &= c(r,r') \int_0^\infty i(\tau) \frac{I(t,\tau,r')}{N(t,r')}\, d\tau. \end{aligned} \tag{5.7}$$

Here $k(t,r,r')$ is the probability of being infected by a partner of risk r'. The partnership function $p(t,r,r')$ defines the rate a person of risk r forms a sexual partnership with a person of risk r'. For random partner choice, this rate is the product of the rate of partnership formation, r, and the fraction of available partnerships that are with people of risk r', $F(r,r')$:

$$\begin{aligned} P_{\text{random}}(t,r,r') &= rF_{\text{random}}(r,r'), \\ F_{\text{random}}(r,r') &= r'N(t,r')\big[\langle rN(t)\rangle\big]^{-1}. \end{aligned} \tag{5.8}$$

To account for partnership biasing, $F_{\text{random}}(r,r')$ is determined by the fraction of partnerships from r' that are both available and acceptable. Thus, if partners of risk r' are accepted by people with risk r with a frequency $f(r,r')$ and partners of risk r are accepted by people with risk r' with frequency $f(r',r)$, then the fraction of partnerships available and acceptable to a person of risk r is

$$F(r,r') = f(r',r) f(r,r')\, r' N(t,r') \left[\int_0^\infty r'' f(r,r'') f(r'',r)\, N(t,r'')\, dr''\right]^{-1}.$$

There is, however, a constraint on $p(t,r,r')$: the total rate that $r-r'$ partnerships form, $N(t,r)p(t,r,r')$, must be equal to the total rate that $r'-r$ partnerships form. In addition, $F(r,r')$ must be discounted to take account of the partner choice that a person from r' has at that time. As an approximation (which may not ensure that people from r have exactly r partners/year), we take an average, and let

$$p(t,r,r') = \frac{rf(r',r) f(r,r')\, r' N(t,r')}{\left(\int_0^\infty r'' N(t,r'') \tfrac{1}{2}\big[f(r,r'')f(r'',r) + f(r',r'')f(r'',r')\big]\, dr''\right)}. \tag{5.9}$$

Substituting Equation (5.9) into Equation (5.7) defines $\lambda(t,r)$.

The system in Equation (5.4) with this choice of $\lambda(t, r)$ gives a model that allows the implications of a wide variety of partner-selection mechanisms to be investigated. Different acceptance functions $f(r, r')$ and contact functions $c(r, r')$ can model different social behaviors and forecast quite different futures for the epidemic. For example, $f(r, r') = 1$ implies random partner selection, and $f(r, r') = 0$ for $r \neq r'$ and 1 for $r = r'$ implies a person and his or her partner always has the same number of partners.

Diffusion-Risk Model. We want to consider the effects of a strong selection preference toward partners from similar risk groups, with more-active people less discriminating than less-active people. As a first step in this direction, suppose that partners are chosen within $r \pm \frac{1}{4} r\epsilon^{1/2}$, according to a Gaussian partnership acceptance function

$$f(r, r') = e^{-(r-r')^2/8\epsilon r^2}. \tag{5.10}$$

In Section VII.B, we compare calculations with this choice of $f(r, r')$ with the random-mixing model, when variations in τ are ignored [$i(\tau)$, $\gamma(\tau)$, and $\delta(\tau)$ are constant]. If we look at the limit as this acceptable range gets small ($\epsilon \to 0$), and keep only the first correction in ϵ, we obtain the diffusion expression in Equation (5.6) for $\lambda(t, r)$ for all $r \ll \epsilon^{-1/2}$ (see [11] for more details).

If we consider only the initial few years of the epidemic, when few AIDS cases have yet occurred, birth and death processes can be ignored and we can assume $\mu \simeq 0$, $\gamma(\tau) \simeq 0$. Approximating the distribution $N(t, r) = N(r)$ by $N_0/(2\bar{r} + r)^4$ as in Figure 10 (Section VI), neglecting variations in infectivity [$i(\tau) = i$], and taking a single contact per partner [$c(r, r') = 1$], then Equation (5.4), with $\lambda(t, r)$ given by Equation (5.6), reduces to a simple diffusion equation for the fraction in each group infected:

$$\frac{\partial v(t,r)}{\partial t} = i[1 - v(t,r)] \left\{ rv(t,r) + \epsilon r^4 \left[5 \frac{\partial}{\partial r} \left(\frac{v(t,r)}{r^2} \right) + r \frac{\partial^2}{\partial r^2} \left(\frac{v(t,r)}{r^2} \right) \right] \right\}, \tag{5.11}$$

where

$$v(t, r) = I(t, r) N^{-1}(r).$$

There are solutions of this diffusion equation (5.11) that have the form $v(t, r) = V(rt)$ (with any arbitrary time shift, $t \to t + t_0$, allowed), and numerical simulations show these similarity solutions are strongly attracting (Section VII.B). Note that considering only the fraction infected in each

group, $v(t, r)$, can give a misleading picture of where the epidemic is spreading because there are so many more people with lower-risk behavior. That is, a small fraction of infected low-risk individuals may be greater in absolute numbers than a large fraction of infected high-risk individuals.

Neither totally random choice nor biased choice only from neighboring risk groups captures the true behavior of people. In the absence of data, however, it is worthwhile to postulate these two extremes and compare the epidemics that each predicts, but we must also look at mixtures of the two behaviors. In the simulations presented in the next section, there is an enormous difference between the two extremes. Sattenspiel (personal communication) proposed that a simple way to look at mixtures of the two behaviors is to take a linear combination of random plus self-preference. Jacquez et al. [31] have used this idea to examine the transition from pure random selection to pure self-selection using a model with four discrete activity levels. They see a large difference in epidemic growth rates, the time to spread across the different activity groups, and the endemic state when the pure self-selection term dominates (over 90%).

In sexually active heterosexual communities, there may be a very different mixing pattern from the ones described here. Also, differences between male and female mixing patterns must be assessed. Data on clients of prostitutes should be gathered and examined to understand not only the activity levels of these clients, but also what their nonprostitute partners are like.

Even within the male homosexual and the IV-needle-sharing communities, behavior patterns are not this simple. Behavior changes over time, and people with many partners one year may have only a few the next, or vice versa. Social groups within which mixing is strong, and between which it is weak, may cause low-activity people in one group to be infected before high-activity people in another group.

The social/nonsocial mixing behaviors modeled by Sattenspiel [51] and Sattenspiel and Simon [52] may also play an important role in the spread of this disease. Models with a variety of mixing assumptions need to be developed and compared, both with each other and with behaviorial and serological studies, to ascertain what complexities are really necessary for modeling HIV spread and which are not.

C. REDUCTION OF THE RANDOM-CHOICE RISK-GROUP MODEL

Under the assumption that migrations and the natural death rate are small and that the contacts between individuals go as $c(r, r') = 1 + h(r)h(r')$, where $h(r) \to 0$ as $r \to \infty$, the system in Equations (5.4) and (5.5) can be simplified by analytically calculating the distribution of infecteds in risk and eliminating the risk coordinate. Besides being faster to solve numerically, the simpler system has the advantage that it is accurate in the r-direction and discretization errors in r are eliminated. To derive the reduced equations, we

first define two functions of time, y and z, which satisfy the equations

$$\frac{dy(t)}{dt} = \frac{\int_0^\infty i(\tau)\int_0^\infty rI(t,\tau,r)\,dr\,d\tau}{\int_0^\infty rS(t,r)\,dr + \int_0^\infty\int_0^\infty rI(t,\tau,r)\,dr\,d\tau}, \qquad y(0)=0,$$
$$\frac{dz(t)}{dt} = \frac{\int_0^\infty i(\tau)\int_0^\infty rh(r)\,I(t,\tau,r)\,dr\,d\tau}{\int_0^\infty rS(t,r)\,dr + \int_0^\infty\int_0^\infty rI(t,\tau,r)\,dr\,d\tau}, \qquad z(0)=0, \tag{5.12}$$

and note that

$$\lambda(t,r) = r\frac{dy(t)}{dt} + rh(r)\frac{dz(t)}{dt}. \tag{5.13}$$

If we assume

$$S(t,r) = S(0,r)\,e^{-ry(t)-rh(r)z(t)} \tag{5.14}$$

and differentiate with respect to t, we recover the first equation in the system (5.4) with μ set to zero:

$$\frac{\partial S(t,r)}{\partial t} = -\lambda(t,r)\,S(t,r). \tag{5.15}$$

Multiplying this expression for $S(t,r)$ by r and integrating over all r, we get one piece of the right-hand side of the equations for $y(t)$ and $z(t)$ in terms of y and z, which we call $g(y,z)$. If we let G be the total susceptible population, then

$$G(y,z) = \int_0^\infty S(0,r)\,e^{-ry-rh(r)z}\,dr, \tag{5.16}$$

and we see that

$$g(y,z) = \frac{-\partial G(y,z)}{\partial y}.$$

Multiplying the equations for $I(t,0,r)$ and $I(t,\tau,r)$ in the system (5.4), (5.5) by r and by $rh(r)$ and then integrating over all r, we obtain equations for the other pieces of these right-hand sides. If we define

$$x(t,\tau) = \int_0^\infty rI(t,\tau,r)\,dr$$

and

$$u(t,\tau) = \int_0^\infty rh(r)\, I(t,\tau,r)\, dr,$$

then

$$x(t,0) = -\frac{dg}{dt},$$

$$u(t,0) = -\frac{d}{dt}\left(\frac{\partial G}{\partial z}\right), \tag{5.17a}$$

$$\frac{\partial x(t,\tau)}{\partial t} + \frac{\partial x(t,\tau)}{\partial \tau} = -\gamma(\tau)\, x(t,\tau),$$

and

$$\frac{\partial u(t,\tau)}{\partial t} + \frac{\partial u(t,\tau)}{\partial \tau} = -\gamma(\tau)\, u(t,\tau). \tag{5.17b}$$

The equations for $y(t)$ and $z(t)$ can then be rewritten as

$$\frac{dy(t)}{dt} = \frac{\int_0^\infty i(\tau)\, x(t,\tau)\, d\tau}{\int_0^\infty x(t,\tau)\, d\tau + g(y,z)}, \qquad y(0) = 0, \tag{5.18a}$$

$$\frac{dz(t)}{dt} = \frac{\int_0^\infty i(\tau)\, u(t,\tau)\, d\tau}{\int_0^\infty x(t,r)\, dt + g(y,z)}, \qquad z(0) = 0. \tag{5.18b}$$

Initial conditions on $x(0,\tau)$ and $u(0,\tau)$ come from the initial infected distribution and the definition of $x(t,\tau)$ and $u(t,\tau)$. The numbers of susceptibles, infecteds, and AIDS cases satisfying Equations (5.4) and (5.5) can then be recovered from the solutions of this simpler reduced system.

In the next section we discuss parameter choices for the models presented in this section. Numerical results are presented in Section VII.

VI. MODEL PARAMETERS

The models discussed in the previous section contain a number of parameters that must be estimated in order to make calculations. Some of these parameters can be estimated fairly well [μ, γ, or $\delta(\tau)$], but for most of them only partial information is known. We wish to explore the effects of parameter changes, within plausible ranges, on the solution of these models. In this section we discuss the information that is known about these parameters and the possibilities that we explore in the numerical simulations presented in the next section.

A. RATE OF DEVELOPING AIDS

The fraction of infecteds developing AIDS within time τ since infection has been estimated for the first 90 months for both the San Francisco hepatitis B cohort [27] and the Hersey hemophiliac cohort [20]. These estimates show that small numbers of adults begin developing AIDS 2 years after infection, with a larger and larger fraction developing AIDS each year up to the end of these studies, when 30% and 25% of these cohorts, respectively, had developed AIDS after about 7 years. Progression-time distributions have been estimated for shorter times for other cohorts.

Unfortunately, error bars are large on all of these estimates because of the small sample sizes. Also, in most cohorts, conversion times are known only within some general time period, with the earlier conversion times the least well measured. In addition, the rate of developing AIDS depends upon the age, health, and sex of an individual as well as the course of the disease, e.g., KS, PCP, or dementia.

We cannot wait another 10 years or more for the data before estimating the distribution beyond 8 years. One way to make these estimates is to choose a reasonable functional form and fit the parameters to the initial portion of the curve using existing data. A reasonable function should have an initial shape similar to the data and should be nonsymmetrical, with some people developing AIDS many, many years after infection. These restrictions still leave the future shape of the curve arbitrary. Weibull, gamma, and log-logistic distributions have been used in various studies by previous authors [37, 42]. We have chosen to use the Weibull distribution of Medley et al. [42],

$$C'(\tau) = pq^p\tau^{p-1}e^{-(q\tau)^p}, \tag{6.1}$$

with $p = 2.4$, $q = 0.11$ for the times from infection to AIDS, primarily because it agrees well with the first 7 years of estimates from the portion of the San Francisco hepatitis B cohort for whom the date of infection can be estimated (George Lemp, personal communication). This distribution, shown in Figure 8, has a maximum at 7.5 years and a median value of 8 years and is chosen such that all infected persons eventually get AIDS. If less than 100% of the infected people get AIDS, the tail of the distribution should be reduced, but the first 7 years should be left unchanged.

To derive the rate $\gamma(\tau)$ of getting AIDS at time τ after infection from $C'(\tau)$, we note that the solution to $I_t + I_\tau = -\gamma(\tau)I(t,\tau)$ is $I(t,\tau) = \exp[-\int_0^\tau \gamma(\tau)\,d\tau]I(t-\tau,0)$. Thus, the fraction of infecteds infected at $t-\tau$ who have not developed AIDS by time t is the exponential coefficient, $\exp[-\int_0^\tau \gamma(\tau)\,d\tau]$. This fraction is also $1-\int_0^\tau C'(\tau)\,d\tau$. Equating these two expressions, we conclude that

$$\gamma(\tau) = C'(\tau)[1-C(\tau)]^{-1}, \qquad C(\tau) = \int_0^\tau C'(\tau_\alpha)\,d\tau_\alpha, \tag{6.2}$$

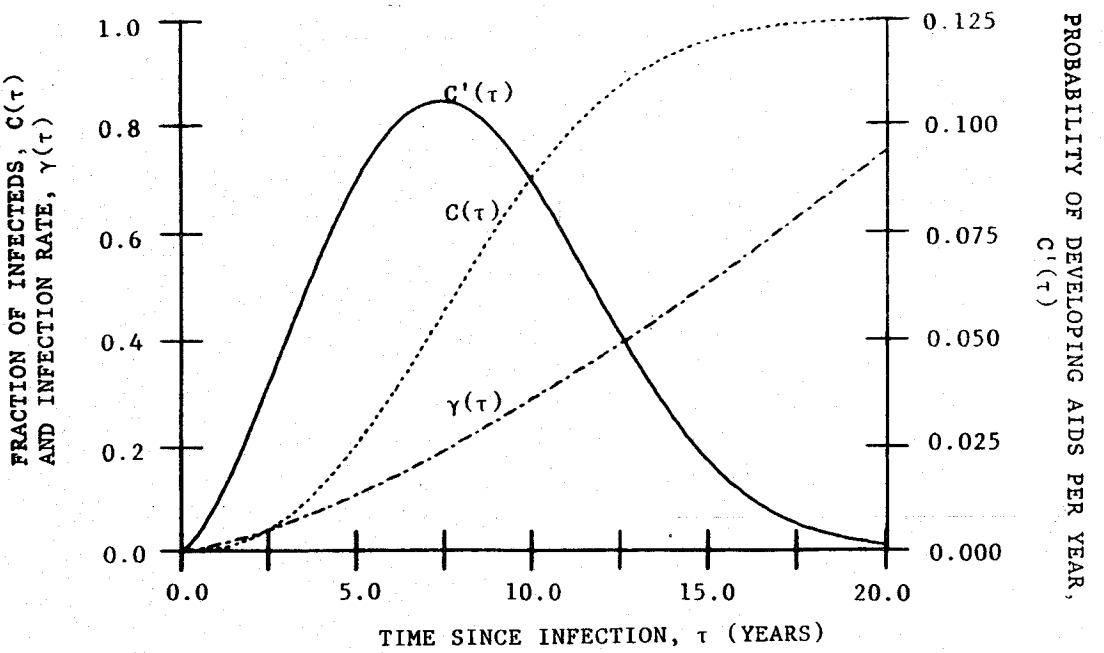


FIG. 8. Conversion from infection to AIDS as given by Eq. (6.1) with $p = 2.4$, $q = 0.11$. Here $C(\tau)$ is the probability of developing AIDS by τ years after infection, $C'(\tau)$ is the probability density of developing AIDS at τ years after infection, and $\gamma(\tau)$ is the rate of developing AIDS at time τ after infection.

is the rate that an individual develops AIDS. $\gamma(\tau)$ is shown in Figure 8 for the Weibull distribution of Equation (6.1).

A possible way to estimate the distribution beyond current conversion times is to collect data on serological markers such as T-4 cell counts and T-4/T-8 ratios as a function of time since infection. These markers indicate the rate of disease progression, even in otherwise asymptomatic individuals; Brodt et al. [8] and Redfield [47] found that over 80% of their cohorts deteriorated in 2- to 3-year periods. If distribution functions for these markers were estimated for different times since infection, then they could be projected into the future to predict the progression-time distribution.

B. *INFECTIVITY*

In Section III.B we postulated a dependence of the infectiousness of a contact on the clinical status of the infector, which is shown as the dotted line in Figure 9 for a time to AIDS of 8 years. This postulate is based on a few measurements of viral presence as a function of clinical status and on speculations about viral interactions with the immune system. Information about actual variations of infectivity with disease progression are anecdotal at this point. Even information on average per-contact infectivity is only good enough to make estimates on its order of magnitude. Padian et al. [45] have used partner studies to estimate an average per-contact infectivity from

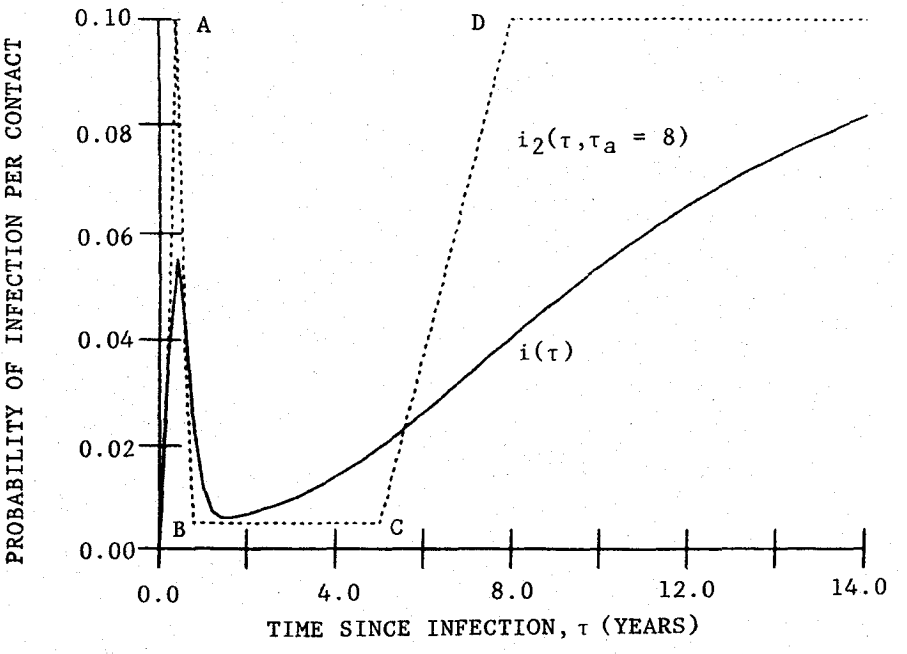


FIG. 9. The infectivity of an average person infected at time τ is a smeared version of the infectivity of an individual. We have postulated an individual infectivity $i_i(\tau, \tau_a) = i_2(\tau, \tau_a) = i^*(\tau/\tau_a)$. The dotted line shows $i_2(\tau, \tau_a)$ for $\tau_a = 8$ years, and the solid line shows the average infectivity $i(\tau)$ given by Equation (6.3) with $C'(\tau)$ as in Figure 8.

man to woman of 0.001 when no other venereal diseases are present. Grant et al. [24] have used seroprevalence estimates to estimate a per-partner infectivity for man-to-man transmission (with receptive and insertive intercourse) of $i_p = 0.10$, but they had no information on numbers of contacts between partners. They also make some estimates for per-contact infectivity, assuming a fixed number of contacts per month, and get a range of 0.004 for 8 contacts to 0.03 for 1 contact per month. Only a study with information about the number of contacts between partners and the clinical status of the partner can give actual numbers, but these data indicate that the average infectivity of a sexual contact probably lies between 0.001 and 0.03.

We assumed above that the infectiousness of a single contact, $i(\tau)$, is the average for all infected adults. The infectiousness of any single individual, $i_i(\tau)$, may have occasional ups and downs as health varies, and these variations will be smoothed out when averages are taken. More than this, there is a wide spread in the rate at which immune systems deteriorate. We can think of $i_i(\tau)$ as the sum of two functions: $i_1(\tau)$, which gives the initial immune response as viral counts first go up and then are depleted by antibody response; and $i_2(\tau, \tau_a)$, which gives the long-term immune response in terms of the individual's time to AIDS, τ_a, after infection. If the time to AIDS is given by a probability distribution as $C'(\tau_a)$, then comparison of a model with τ_a explicit and our model without τ_a shows that the

average infectiousness is

$$i(\tau) = i_1(\tau) + \int_{\tau}^{\infty} i_2(\tau, \tau_a) C'(\tau_a)\, d\tau_a [1 - C''(\tau)]^{-1}. \qquad (6.3)$$

Figure 9 shows the effect of this convolution on a speculated $i_i(\tau)$.

Estimates of the time between infection and antibody response are difficult to make. Not only are accidents to health-care workers with documented seroconversion rare, but also few people at risk have been tested frequently enough to obtain good estimates of their seroconversion dates. Thus, this time interval may be from a few weeks to a few months and may be different for different individuals. The relation between viral presence and antibody response also is not well established. Thus, the average width of the initial peak and the ratio between the maximum and minimum values are unknown.

For the τ-model calculations of the next section, we have taken $i_1(\tau) = 0$ and $i_2(\tau, \tau_a) = i^*(\tau/\tau_a)$. We use a piecewise linear infectivity, $i^*(\tau/\tau_a)$, as shown in Figure 9. The solid line in Figure 9 shows the effect of applying Equation (6.3) to the Weibull distribution of Figure 8 and the $i^*(\tau/\tau_a)$ shown as $i_2(\tau, 8)$. We investigated the effect on model solutions of changes in this profile for $i^*(x)$.

C. *DEATH RATES*

The death rates μ and $\delta(\tau)$ are the model parameters for which the best data exist. If we take μ^{-1} to be the average lifetime of an adult, it is around 70–80 years. On the other hand, if we want μ to represent the rate of attrition out of the at-risk community, a μ^{-1} of 30–50 years is more reasonable. In our calculations, we use $\mu = 0.02$ years^{-1}.

The probability of death once AIDS symptoms appear can be estimated from CDC mortality data, where deaths are recorded according to diagnosis date. The rate of death is high at first and gradually decreases. An exponentially decreasing probability density $D'(\tau)$ for death as a function of time τ since developing AIDS, which gives a constant death rate $\delta(\tau)$, fits adequately. A slightly better fit is found by taking the density function to be

$$D'(\tau) = d_1 \exp\left[- d_2 \tau (1 + d_3 \tau)^{-1} \right], \qquad (6.4)$$

where τ is the time since AIDS symptoms appear and $d_1 \simeq 1$ is chosen to normalize the area to 1 at $\tau = 20$ years. Now we get the rate of death to be decreasing in τ:

$$\delta(\tau) = D'(\tau)[1 - D(\tau)]^{-1}, \qquad D(\tau) = \int_0^{\tau} D'(\tau_d)\, d\tau_d. \qquad (6.5)$$

$d_2 = 0.075$ and $d_3 = 0.05$ give reasonably good fits to the CDC data, with 48% dead in 1 year and 90% dead about 5 years later. A recent follow-up of AIDS cases found that deaths were severely underreported [26]. Thus, this distribution might underestimate the true death rate due to AIDS. This underestimate will be somewhat less severe than it might have been because of the widespread use of AZT.

D. DISTRIBUTION OF RISKS

Sexual-activity data from studies of homosexual men show that there is an enormous variation between individuals in the numbers of partners and the amount and type of contacts. Participants in the Multicenter AIDS Cohort Study (MACS), who were questioned between April 1984 and March 1985, reported between 1 and 500 male partners in the previous 6 months, with a mean of between 5 and 10 [34]. The San Francisco Men's Health Study recorded the numbers of their respondents according to the groupings 0,1,2–9,10–49, $\geqslant 50$ partners in the 2 years before June 1984 [59]. Homosexual men surveyed in 1984 in London and grouped according to 0, 1–5, 6–50, 51–00, and $\geqslant 101$ partners in the previous year show a similar amount of variation (data from T. McManus reported in [40]).

These data are available only in interval form, whereas we use a continuous distribution in our model. Unavailability of continuous distributions is a common problem with data. To derive estimates of this continuous distribution, we first formed the cumulative data set of how many people had no more than $x_1, x_2, \ldots$ partners in the time interval, where $x_1, x_2, \ldots$ are the top values for the interval. The last x_i is chosen somewhat arbitrarily. We then interpolate the cumulative data with a smooth, monotonicity-preserving interpolant, such as constrained cubic splines. Differentiating the interpolant gives the continuous density function. Data sets from different studies, with different intervals, can be compared, or combined, using weighted linear combinations of the interpolants, with weightings appropriate to sample size or other knowledge (such as date of sample or sampling procedure). Figure 10 shows the continuous density function obtained from combining the McManus data with data from Carne and Weller, also reported in [40].

The density functions from these interpolations of the San Francisco and London data can be used to derive average partnership densities and variances for each grouping. These estimates (especially the variance) depend on how large the maximum was assumed to be for the group with $\geqslant 50$ or $\geqslant 101$ partners. The MACS study indicates that this number is large because there are people far out in the distribution, giving it a long tail [34]. For the London data, which are given in terms of partners/year, a simple function that approximates the data is $0.06(1+0.02r)^{-4}$. This function has a mean of 25 partners/year, matched to the mean of the interpolant, and a variance of $25\sqrt{3}$ partners/year. At $r = 75$, the function is 0.0015.

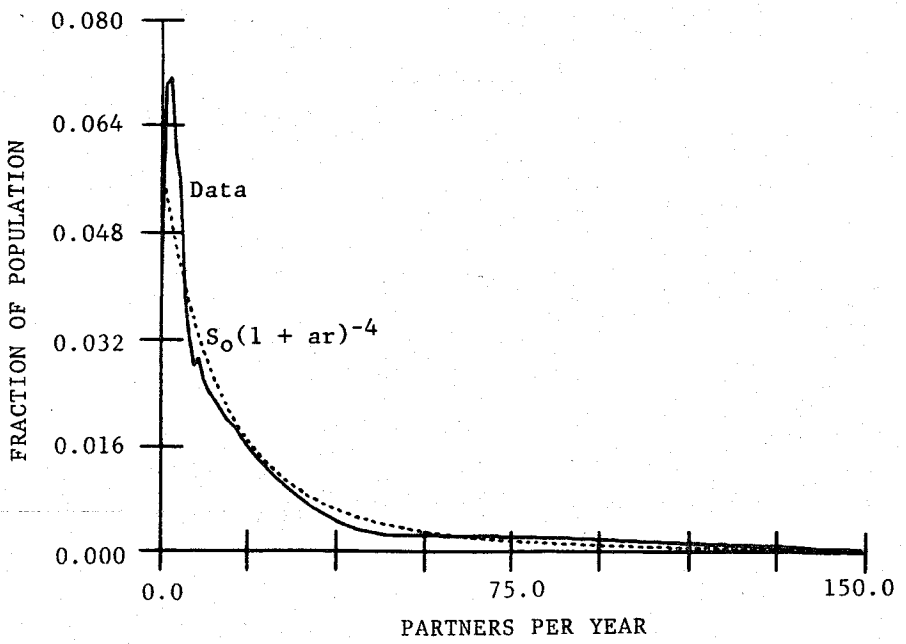


FIG. 10. The distribution of homosexual men attending STD clinics in London, obtained by combining data of McManus (partners/year) and of Carne and Weller (partners/month) using the procedure outlined in the text. The dotted line shows the inverse quartic with the same mean as the data. (Data reported in [40].)

The fact that this inverse quartic function looks much like the data is interesting. For functions of the form $S_0(n) = a(1+br)^{-n}$, the exponent $n = 4$ is the smallest integer choice before the variance becomes infinite. If we expect humanity to be as variable (within finite reason) as possible, we might predict that all risk behavior will follow a similar distribution with n between 3 and 4. The dotted line in Figure 10 shows the inverse quartic, with the mean matched to the data. In the risk-model calculations of Section VII, we used this inverse quartic, with a mean of 24 partners per year.

Information on the number of contacts between different types of partners (long-term, casual, prostitutes) is scarce, even for these homosexual cohorts. This critical information is beginning to be collected [32]. Because transmissibility through different types of contacts may be different, the frequency of each type of contact needs to be quantified. Without such knowledge, the best that we can do is to make some reasonable assumptions and explore various possibilities.

The assumptions that we use are that people with large numbers of partners have one contact with each partner and that people have more contacts with each partner when both partners have fewer partners, up to some maximum number. For simplicity, we use the contact function $c(r, r') = 1 + (c_1 - 1)\exp[-c_2(r + r')]$ and vary the constants c_1 and c_2.

Behavior in the homosexual community has changed substantially since these responses were recorded. By mid-1982, the first news stories on AIDS

began to have an impact [2]. The change in homosexual behavior through fewer contacts or safer sexual practices is reflected both by the drop in rectal gonorrhea in San Francisco [33, 23] and in the results of cohort surveys [59, 38]. We would eventually like to incorporate these changes, but we can first use our model to ascertain whether it captures the infection pattern that occurred before this change. Perhaps information from contact-tracing studies [6] can be used to understand the important questions of partner selection and frequency of various types of contacts between partners.

Similar information is needed about heterosexual behavior and about needle use. Who does what with whom and how often are very important questions to answer if we are to understand this pandemic.

VII. SAMPLE CALCULATIONS

Our focus has primarily been on the qualitative features of the early growth of the epidemic. Therefore, the calculations in this section compare the effects on the growth of the infected population as parameters are varied. We compare the τ-model and three-risk based models (the random-mixing, the biased-mixing, and the diffusion model) with no τ-dependence. In the τ-model, we examine the importance of initial conditions and of the time variation of $i(\tau)$. For the risk-based models, we examine the number of infecteds versus risk and show that there are substantial differences in predictions for the growth of the epidemic. Also, there are significant differences in who is being infected in the random-choice and biased-partner models.

We have focused on early growth because it is important to understand how the epidemic moves into new populations and which interactions are important in its transient dynamics. Understanding these transient dynamics is the only way to understand which new populations are at risk and what the short-term effects of behavior changes and medical advances will be. We are still in the early stages of this epidemic, so the data that we have come from these stages. We emphasize that these models are too simplistic to give accurate predictions of the AIDS epidemic and that the following calculations are meant only to illustrate the behavior of the models.

A. τ-MODEL

We first calculate the solution of the model in Equation (5.3), using (6.1), (6.4), the parameter values described in Section VI, and the initial conditions

$$\begin{aligned} S(0) &= 10, \\ I(0,\tau) &= p^{-1}523.8(\tau_0-\tau)^2[1-C(\tau)]\times 10^{-6}, \qquad \tau \leqslant \tau_0, \\ A(0,\tau) &= 523.8(\tau_1-\tau)^2[1-D(\tau)]\times 10^{-6}, \qquad \tau \leqslant \tau_1. \end{aligned} \tag{7.1}$$

$I(0,\tau)$ and $A(0,\tau)$ are zero for $\tau \geqslant \tau_0$ and $\tau \geqslant \tau_1$, respectively. The units are millions of people and years. The scalar parameters used in the first set of calculations were $\mu = 0.02$ year^{-1}, $r = 36$ partners/year, $p = 1$, $\tau_0 = 1.8$ years, and $\tau_1 = 0$ years. Equations (6.2) and (6.5) were used for the rates of progression from infected to AIDS and from AIDS to death. The individual infectivity $i_2(\tau, \tau_a) = i^*(\tau/\tau_a)$ in Equation (6.3) was a piecewise linear approximation $L[(\tau_1, i_1),(\tau_2, i_2),\ldots]$, shown as the dotted line in Figure 9, which for $\tau_a = 8$ years connects the (τ, i) data points

$$i_i(\tau, \tau_\alpha) = L[(0,0),(0.1,0),(0.4,0.01),(0.7,0.005),(5.0,0.005),(8.0,0.01)]. \quad (7.2)$$

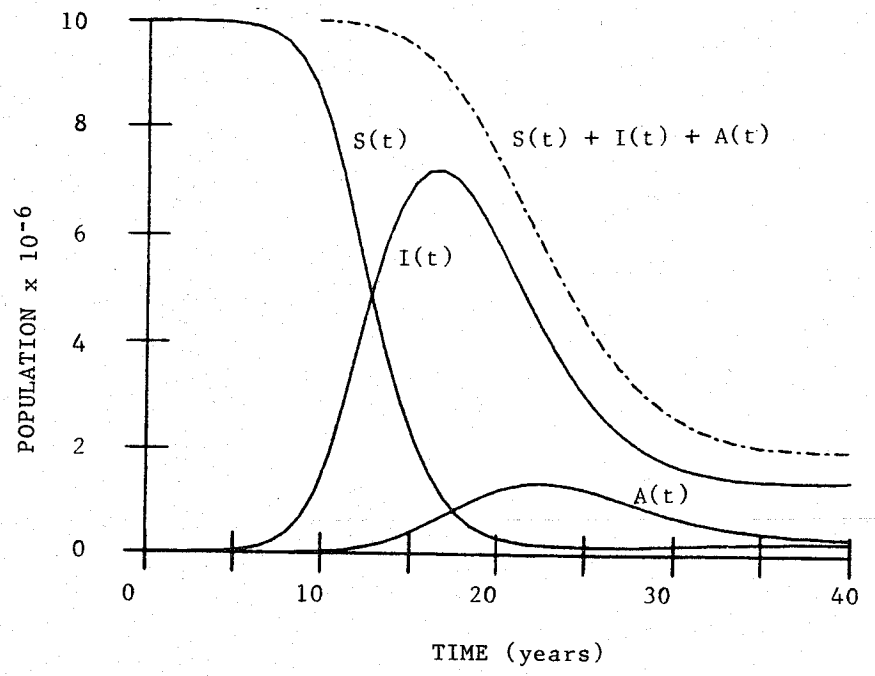
This distribution and the resulting $i(\tau)$ are shown in Figure 9. We also examine the effect of varying some of these data points.

The solutions were integrated in time with an explicit Adams-Bashford-Moulton method to an accuracy of 10^{-6} per unit time. The τ-derivatives were calculated with fourth-order finite differences on a uniform grid of 201 mesh points. The grid spacing and error tolerance were varied to check convergence of the solutions.

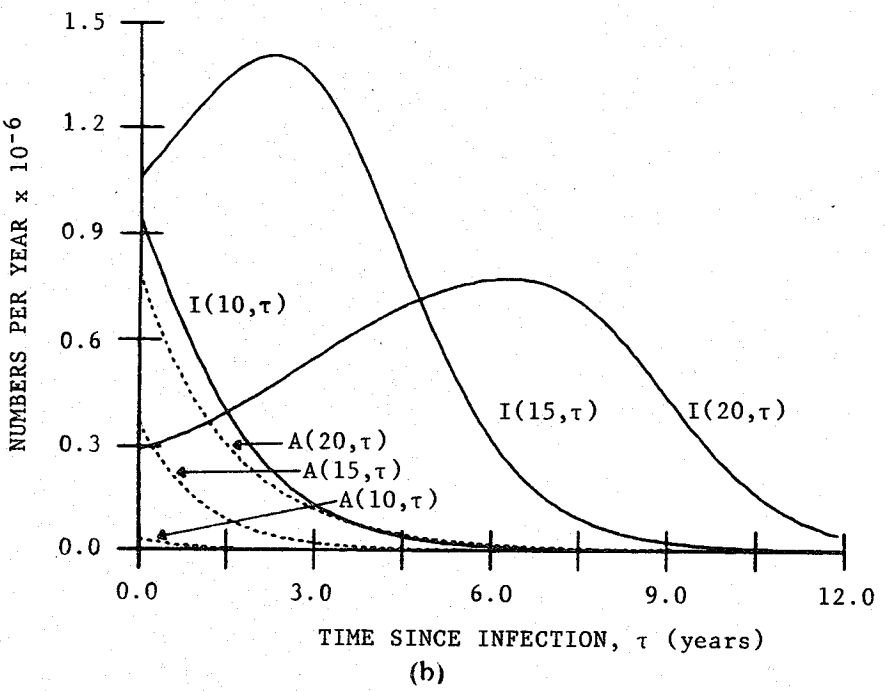
The solution in Figure 11(a) illustrates how the susceptibles steadily decline to near-equilibrium values after 40 years. Initial growth of infecteds and AIDS cases is exponential, unlike AIDS cases in the United States (Section IV). The infection saturates the total population in about 17 years, after which it is greatly reduced because of AIDS deaths. In Figure 11(b) note that the rate of people infected per year at $t = 15$ years has a maximum at $\tau = 3$ years. This maximum moves out and decreases with time because of the depletion of susceptibles [Figure 11(a)].

By using caseload data, probability density functions can be constructed to determine what fractions of the infected population are in each stage of the disease or have developed specific opportunistic infections as a function of time since HIV infection. These distributions can be applied to the predictions of the infected populations, such as the ones in Figure 11(b), to determine how many people will be in each disease stage at any given time. These derived quantities and estimates are a major advantage of calculating the time since infection as a variable in the model.

By varying the infectivity profile, we can dramatically change the rate at which the susceptible population is infected. In Figure 12 we show calculations with four different infectivity profiles. The average infectiousness of an individual, $\int_0^1 i^*(x)\,dx$, is the same for all four profiles. When the amount of infectiousness in the initial peak is modified, the center region is raised or lowered accordingly. The difference in the transient solutions illustrates that the shape of the initial peak in infectivity is important. The shape is important because more people were infected recently (low τ) than 5–7 years



(a)



(b)

FIG. 11. (a) The solution of the model in Equation (5.3) with the initial conditions $\tau_0 = 1.8$ years, $\tau_1 = 0$, $p = 1$, and infectivity as in Equation (7.2). Here $I(t) = \int I(t,\tau)\,d\tau$ and $A(t) = \int A(t,\tau)\,d\tau$. (b) The distribution of infected and AIDS cases during the calculation at times 10, 15, and 20 years.

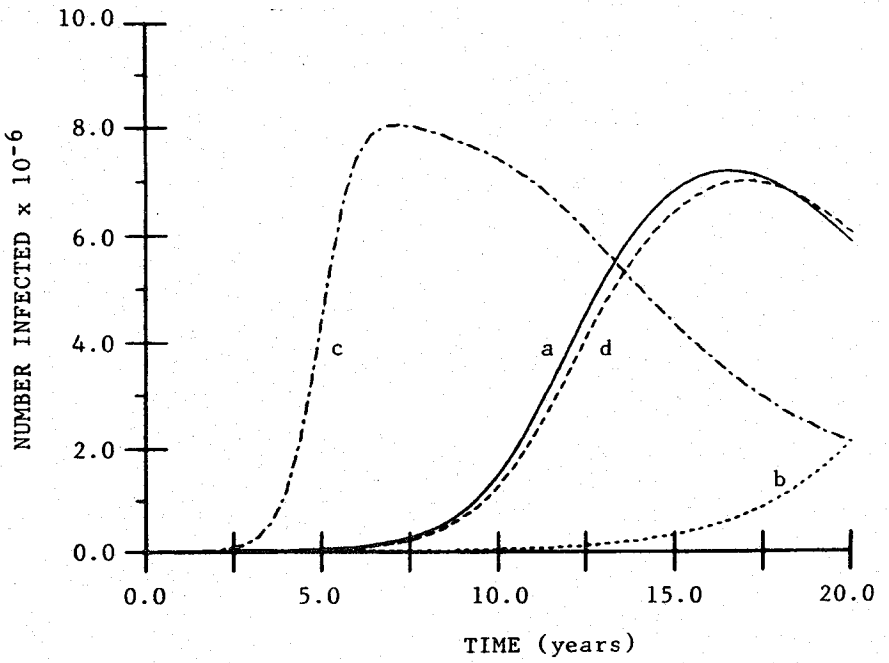
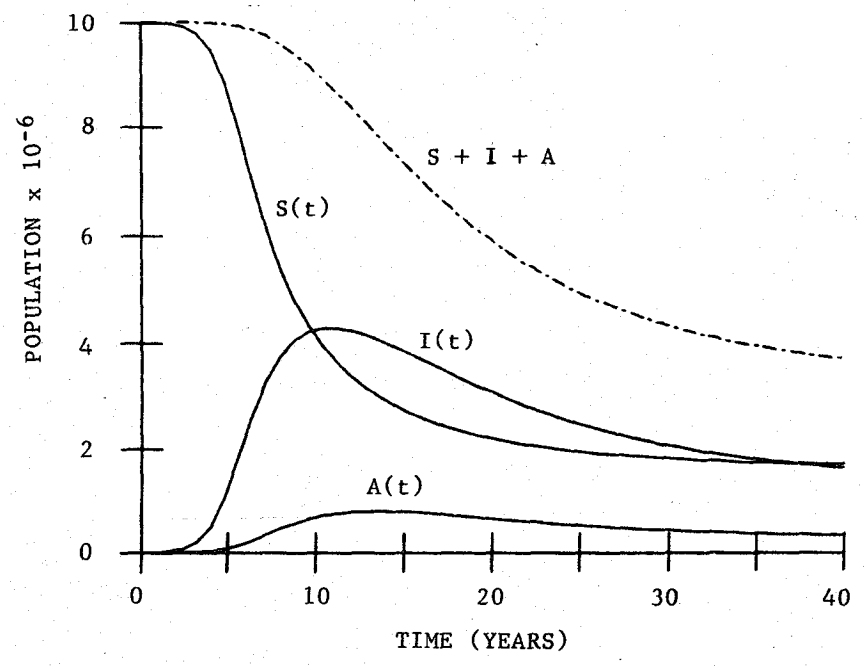


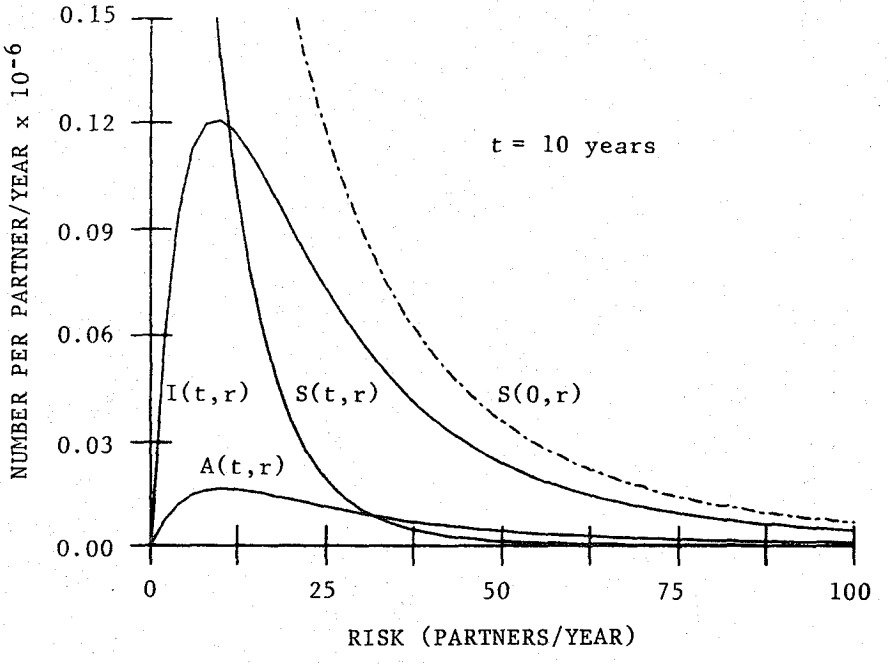
FIG. 12. The rate at which the susceptible population is infected depends upon the infectivity profile even when the area under $i_i(\tau, \tau_a)$ remains unchanged. $I(t)$ was calculated with Equation (7.2) with the following changes: (a) (solid) unchanged; (b) (short dashes) initial peak lowered to $(\tau, i) = (0.04, 0.01)$ and center region raised to (0.7,0.01), (5.0,0.01); (c) (dash-dot) initial peak raised to (0.4,0.2) and center region lowered to (0.7,0.0), (5.0,0.0); (d) (long dashes) final peak raised to (8.0,0.3) and delayed by extending the center region to (7.0,0.005).

ago (high τ). Also, the high infectivity for $\tau > 5$–7 years is important because of its long duration. However, the shape of the late high infectivity is not as important as the shape of the front peak during the early part of the epidemic, because there are relatively few people in this late period. If the infected population remains active in this late, highly infectious period, the epidemic will spread much faster than if they discover they are infected and reduce their sexual activity. Here again we see the need to make testing widely available.

Next the initial conditions of Equation (7.1) were changed to match the current (1988) AIDS case data and the estimates for infecteds from Section IV by setting $p = 0.72$, $\tau_0 = 14.3$ years, and $\tau_1 = 6.8$ years. When we compared these solutions with Figure 11(a), starting at time 8.75 years, we found that the solution from these initial conditions differed for only a couple of years. After 4 years, the calculations were essentially identical and hence are not included here. If $i(\tau)r$ is smaller, the effect of the initial conditions persists longer and imposes cubic growth of the cumulative AIDS cases for the first 5 years. This initial cubic growth would be due only to the past cubic growth of infecteds in the initial conditions of Equation (7.1).

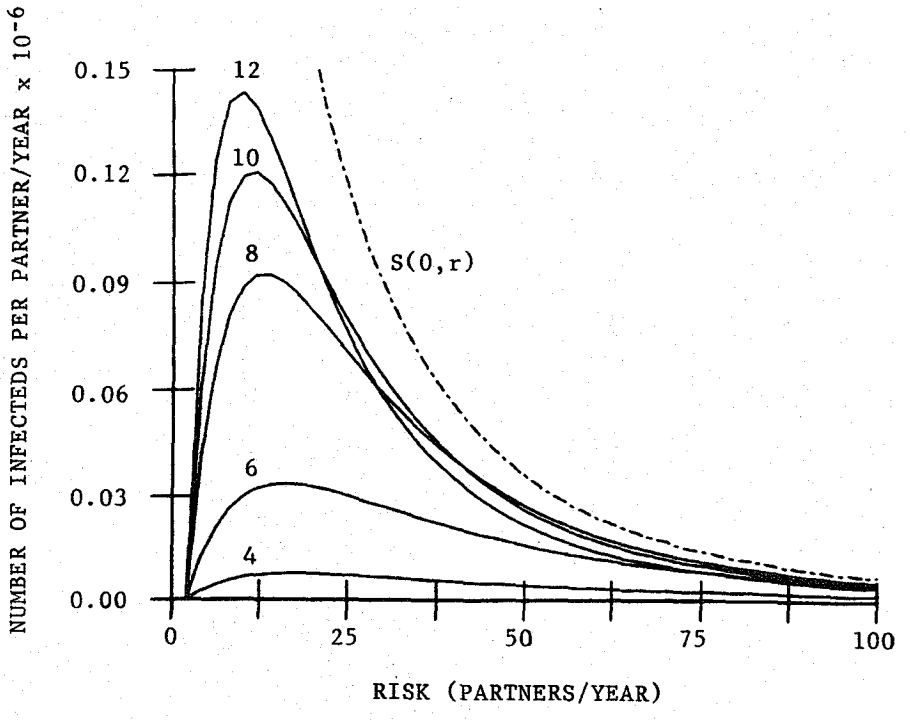


(a)



(b)

FIG. 13. (a) The random-mixing $[c(r, r') = 1]$ risk-based model in Equations (5.4) and (5.5) exhibits an initial exponential growth in infecteds. This behavior is not unlike the models where an average risk behavior is used for all susceptibles. $S(t) = \int S(t, r)\, dr$, $I(t) = \int I(t, r)\, dr$, and $A(t) = \int A(t, r)\, dr$. (b) The majority of those infected and with AIDS at time $t = 10$ years are in relatively low risk groups in the random-mixing model. (c) The low-risk susceptibles are infected early in the epidemic in the random-mixing model. The distribution of infecteds is shown at times 4, 6, 8, 10, and 12 years, indicated on the curves.



(c)

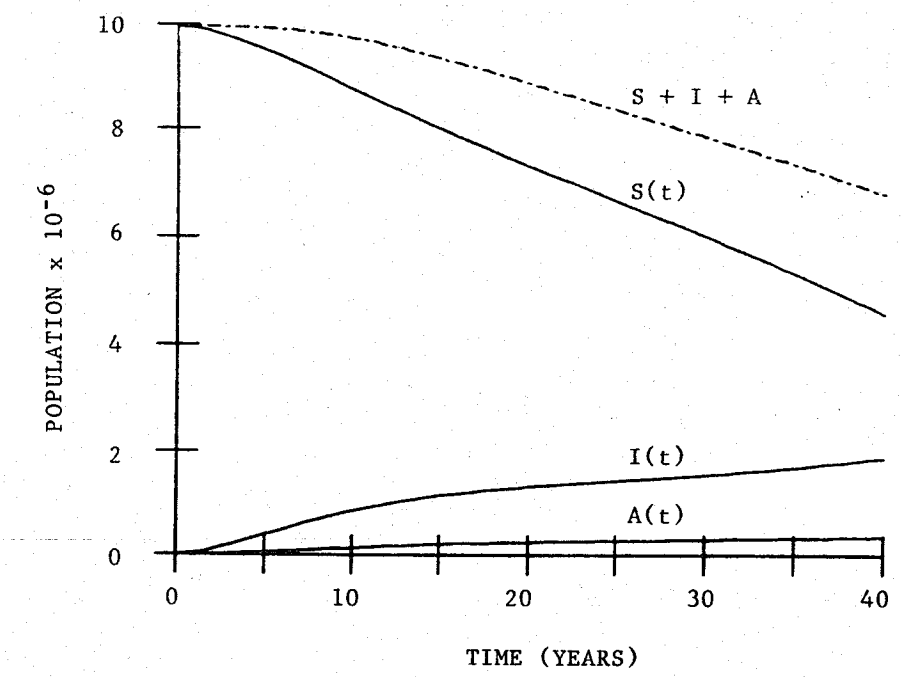
FIG. 13. (continued)

B. *RISK-BASED MODELS*

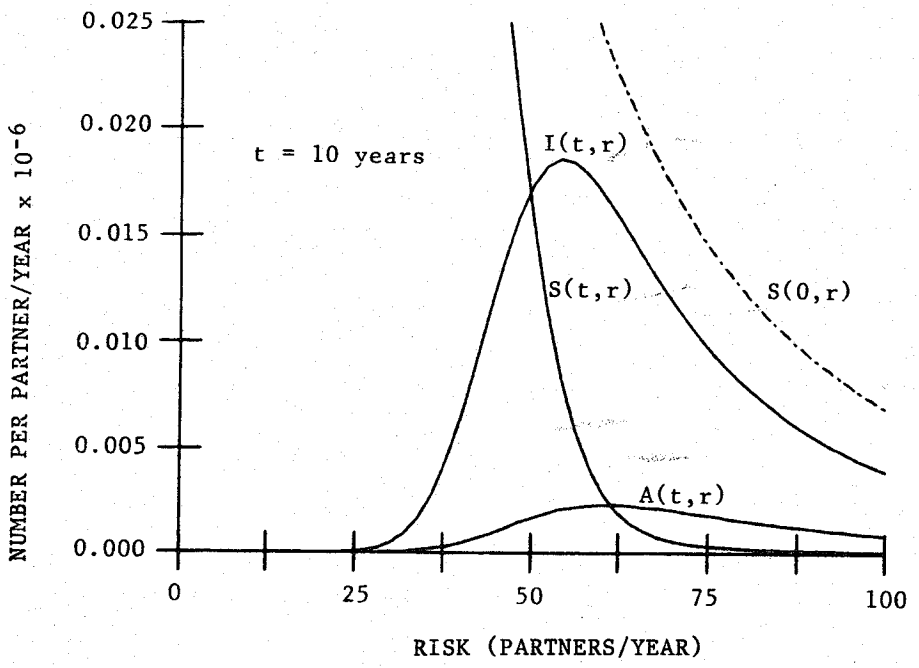
To simplify the calculations and analysis, we eliminated the τ-dependence in the risk-based model (5.4). The τ-independent parameters were defined to be the average values, $\gamma = 0.1$, $\delta = 0.5$, and $i = 0.025$. The initial susceptible population is distributed in risk as an inverse quartic $S(0, r) = S_0\, 3(2m)^3(2m + r)^{-4}$, with total population $\int S(0, r)\, dr = 10$ million, mean $m = \int rS(0, r)\, dr\,(10 \text{ million})^{-1} = 24$ partners/year. There is migration into all risk categories with migration rate equal to the natural death rate, $\mu = 0.02$ years^{-1} times $S_0(r) = S(0, r)$. Initially, there is a Gaussian distribution of 0.001 million infected individuals, centered at risk $r = 175$ partners/year, with height 0.0001 million-years/partner, and no AIDS cases.

Random-Mixing Model. The risk-based calculation shown in Figure 13 used $\lambda(t, r)$ from Eq. (5.5), with the contact function $c(r, r') = 1$. This calculation corresponds to unbiased random mixing across risk groups with a single contact per partner. With $\mu = 0$, this calculation is described by the reduced model described in Section V.C. Because most of the susceptibles have low-risk behavior (small r), a consequence of random partner choice is that most of the partners of high-risk behavior people have low risk. This result is contrary to the sparse sociological data that are available.

Because most of the partners of high-risk people are low-risk, the high-risk group acts as a pool of infection for the lower-risk group, causing



(a)



(b)

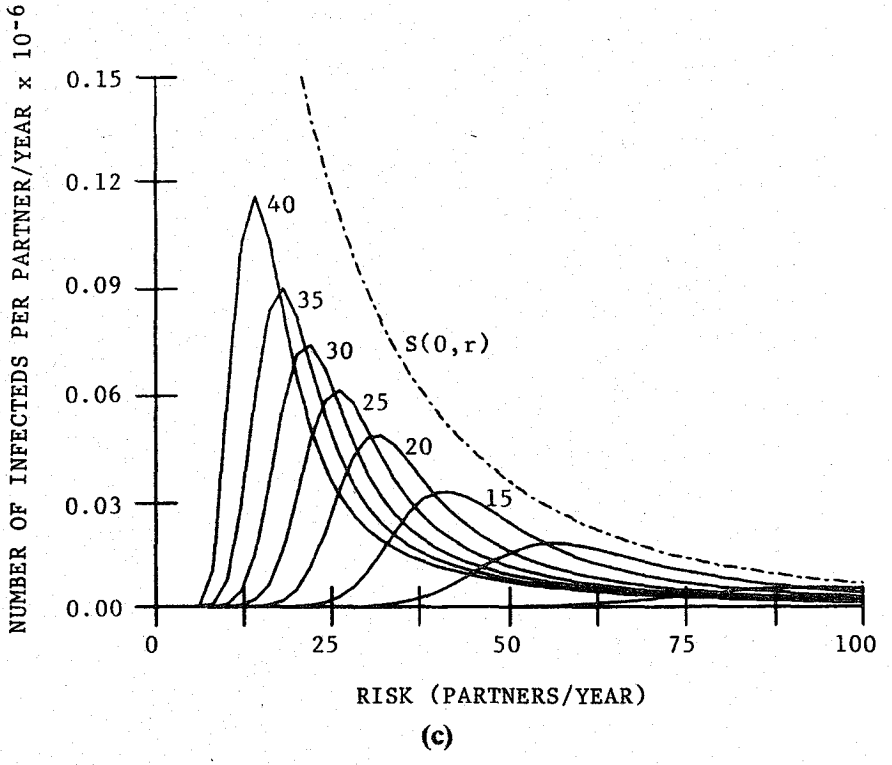
FIG. 14. (a) The infecteds in the biased-mixing model in Equations (5.4), (5.7), (5.9), and (5.10) with the contact function in Equation (7.3) grows as a quadratic polynomial for many years, after a brief transient due to the initial conditions. (b) The AIDS cases at time $t = 10$ years are centered at a substantially higher risk behavior than in the random-mixing model. Note that the AIDS cases have a much flatter distribution in risk than do the infecteds. (c) The infected population forms a wave that sweeps from high-risk behavior groups into lower-risk groups. The distribution of infecteds is shown every 5 years, at the times marked on the curves.

the lower-risk populations to become infected very quickly, with most of the early AIDS cases in lower-risk categories. The distribution of the populations at 10 years is shown in Figure 13(b) as a function of risk. The distribution of the infected population is shown in Figure 13(c) for a sequence of times. We remark that if we had been plotting the fraction of the population that is infected, $I(t,r)/N(t,r)$, then a saturation wave of the fraction of infecteds in a particular risk group would sweep from the high-risk categories into the lower-risk categories. However, because there are so many more susceptibles at low risk than at high risk, the total number infected does not have this shape at all.

The initial growth of infecteds and AIDS cases is exponential for this model, just as for the τ-model. By time 10 years, the infection saturates the population and the susceptibles are greatly reduced. However, the people at lowest risk are protected, so the equilibrium susceptible population is somewhat larger than for the τ-model and the total number infected is somewhat less.

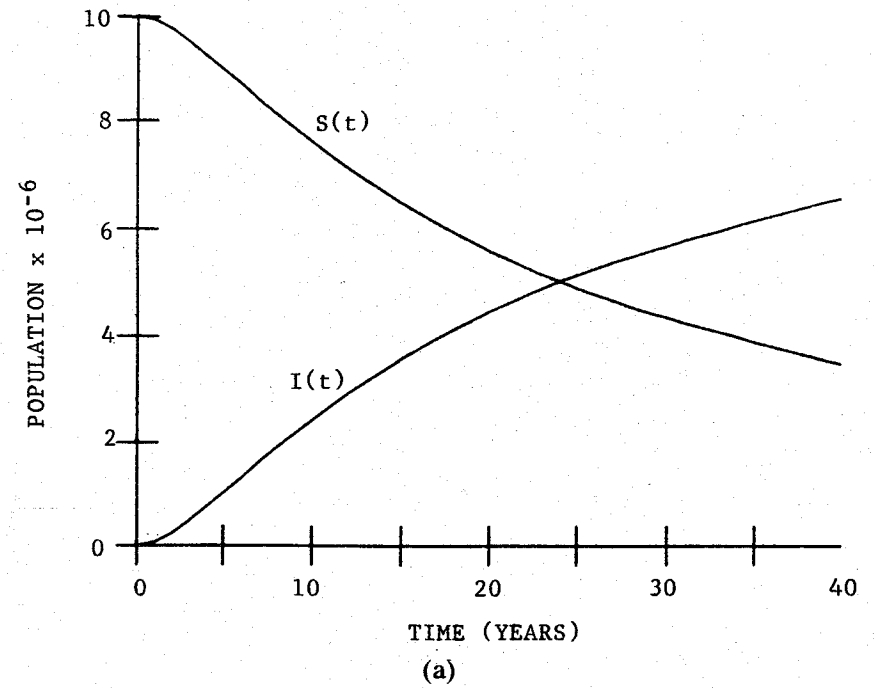
Biased-Mixing Model. Next, we enforce the biased-mixing restriction that people have contacts only with individuals having similar risk behavior. The contact function

$$c(r,r') = 1 + e^{-c_1(r+r')}(c_0 - 1) \tag{7.3}$$

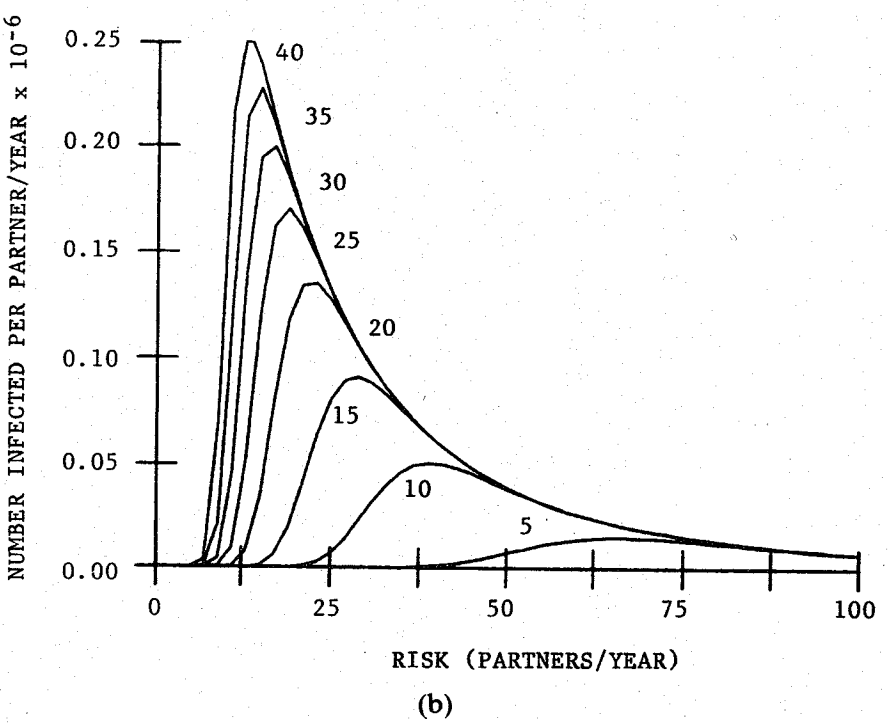


(c)

FIG. 14. (continued)



(a)



(b)

FIG. 15. (a) The solutions of the diffusion model (5.11) are polynomial, rather than exponential. Because infecteds are not removed upon developing AIDS, they grow faster than in Figure 14. Also the susceptibles are infected faster by the larger infected population. (b) The infected saturation wave solution of the diffusion model sweeps into lower-risk behavior similar to the solution in (a).

and the acceptance function $p(r, r')$ defined by Eqs. (5.7), (5.9), and (5.10) are used. In the calculations shown in Figure 14 we used the constants $c_0 = 11$, $c_1 = 0.1$, and $\epsilon = 0.0125$. Note that in contrast to the random-mixing example, the infecteds in Figure 14(a) grow as a quadratic polynomial as the saturation wave shown in Figure 14(c) sweeps from high-risk into lower-risk groups. The growth of infecteds is much slower: after 40 years the lowest-risk susceptibles have not yet become infected. The current number of AIDS cases at 40 years is about the same for the two calculations, however, and the number infected is actually greater. Thus, the eventual impact of the epidemic may be just as devastating, depending on whether it can continue to reach the larger, lower-risk populations or not.

Note in Figure 14(b) that the AIDS cases at time $t = 10$ years have a broader distribution than do the infecteds, and AIDS cases lag behind the infection wave. This scenario is closer to the observed distribution of risk behavior in the early AIDS cases than the unbiased-mixing model. Although this model is far better than the unbiased-random-mixing model, we believe that it will be significantly improved if we add a blend of biasing mixing with a lower level of unbiased random mixing. Also, people do not maintain the same risk behavior forever. Therefore, we are considering adding a mechanism that will allow for some migration of risk behavior.

Diffusion-Risk Model. The solution of the nonlinear diffusion model in Equation (5.11) is shown in Figure 15 for $\epsilon = 0.01$ and the same initial conditions as the previous risk models. The infected saturation wave in Figure 15(b) is similar to the one shown in Figure 14(c), which this equation approximates. The major distinction between the two models is the lack of birth and death processes with the model in Equation (5.11). This lack causes the infected population to be larger because infecteds are never removed by AIDS, and the larger infected population can then infect more susceptibles, causing a faster epidemic. Also, new susceptibles are not created, causing an even more rapid depletion of the population. The epidemic is, however, polynomial in time, as was the calculation of Figure 14.

VIII. SUMMARY

Mathematical models for the spread of the AIDS virus are essential tools for understanding the AIDS epidemic. Using models, we can investigate competing forces and study their interactions to improve our understanding of the relationships between the social and biological mechanisms that influence the spread of the disease. The relative influence of various factors on the spread of the epidemic, as well as the sensitivity to parameter variation, can be ascertained. We can use this knowledge to help set priorities in research. Once the important forces have been identified, we can develop models with which we can run computer experiments comparing the

outcome of different assumptions and strategies for controlling the epidemic. Computer experiments can save time, resources, and lives, allowing us to predict the future and acting as a control group for true experimental situations.

As a first step in developing a reliable model, we have used a simple deterministic model to explore the impact of various plausible shapes for the infectivity as the time since infection increases. These calculations, which use an average-risk behavior, point out the importance of measuring the variability of the infectiousness of an individual during the disease.

We have then used models that stratify the population according to the number of sexual partners per year and have compared random partner choice with a strong bias of "like prefers like." The two mixing patterns result in radically different epidemics. This difference indicates that much more must be known about the interactions between people that lead to AIDS-virus spread before it will be possible to accurately predict the AIDS epidemic. The number of sexual partners that people have, the partner-selection process, and the amount and type of contacts between partners must be understood and correlated with sociological information about the partners, such as how many partners one's partners have. Similarly, patterns of needle sharing by drug users and the effect of this drug abuse on sexual behavior strongly affect this epidemic.

In our analysis, we have focused on the initial growth of the epidemic. If we are to predict where this epidemic is going, we must fully understand its transient dynamics, including the response to changes in the environment of the epidemic. The epidemic will not reach an equilibrium endemic state for a very long time, partly because of the long conversion times from infection to AIDS, during which a person can transmit the virus. This time factor makes AIDS unlike many other epidemics, including measles [15], gonorrhea [28], and syphilis [39].

Another reason AIDS is different is that medical advances and changes in lifestyle will greatly modify the epidemic. Education programs are being launched to promote condom use, having fewer sexual partners, the use of nonoxynol-9, the use of sterile needles, and similar practices. The infectiousness and susceptibility of high-risk individuals in the heterosexual community may be significantly reduced if programs are initiated to quickly identify and treat other STDs. More people are being tested for antibodies to HIV and counseled on the implications of the test results. Treatments are being developed that will prolong the lives of infected persons and perhaps lower their infectivity. A partially effective vaccine may eventually be developed.

Models can be used to investigate the effects of each of these programs on the course of the epidemic only if they can capture the transients of the epidemic.

In developing models, we must also decide on what questions we want to answer. If public-health officials are to attack this epidemic efficiently, then

they need to know which groups of people are most at risk of infection. Models that distinguish between behavioral groups may help predict where the infection is likely to go next. Our risk-based model is aimed toward this question, although it is at present too simplistic to use for this purpose.

We can choose parameters in our preferential-mixing model that ensure that AIDS cases in the numerical simulations match the past history in the United States. Many other reasonable models can also quantitatively fit these cases but may predict a very different future. Quantitatively matching past AIDS cases is not, therefore, sufficient to distinguish between models. Qualitative discrepancies between AIDS cases and any model need to be explained; for example, models with initial exponential growth do not fit the U.S. AIDS case data.

Models must be compared with data from studies connecting risk behavior with infection. For example, we plan to compare our preferential-mixing model with the San Francisco hepatitis B study. In this study of sexually active homosexual men, which started in 1978, information on numbers of steady and nonsteady sexual partners and numbers of contacts per partner was collected, and a series of serum samples were stored from a subset of the men. Many of these samples have been tested for HIV, and so a correlation between sexual behavior and time of infection can be made and compared with our model. Inconsistencies will be seen, and the model will have to be revised to account for them.

Limitations of the data will greatly influence the capability of models to accurately predict the future. Many of the sensitive parameter values, such as the magnitude and variability of infectiousness, will be known accurately only after years of careful study. The current lack of a national AIDS data-base center to collect, analyze, and distribute the available data is a severe block to our understanding. We support establishing a data center that will encourage closer collaborations between modelers and data collectors. The modelers will be more driven to answer questions raised by data, and they will pose questions that will suggest new data that should be collected and more effective sampling strategies to reduce the variance in the results. Focusing on the data helps bridge the gap between mathematical modelers and epidemiologists. Fortunately, the creation of a national AIDS data-base center is one aspect of the AIDS epidemic that can be solved with appropriate funding.

Unlike many other diseases, HIV infections can persist (invisible and seemingly dormant) in a few isolated individuals (with low sexual activity) for long times. This feature can cause sporadic local epidemics whenever the infected individual passes the virus to a highly sexually active person. In these situations the virus can spread rapidly without warning, infecting a great many people. These sporadic events should be modeled by a stochastic rather than a deterministic model, such as ours, that smooths over the sporadic effects of such local random features. Because of the long time

between HIV infection and AIDS, such situations can be ascertained only through vigilant HIV testing and case tracing.

It is important to use models to understand the spreading in parts of the world other than Western Europe and the United States. The current prevalence of HIV infection in central Africa (up to 25% in metropolitan areas) raises serious political and social concerns. Estimates that up to 26% of the adults in some regions in Asia, Africa, and Latin America are annually infected with gonorrhea indicate not only that behavior may be more conducive to the spread of STDs there, but also that the cofactors for AIDS are different. The presence in central Africa of cofactors, such as genital ulcers and the lower general health of the population may be sufficient to explain the rapid heterosexual spread of HIV infection there relative to the United States and Europe; but some aspect of sexual behavior may also be important. Also, condoms and spermicides are used less frequently in these regions than they are in the United States. We need to understand the reasons for regional differences, before we can predict the epidemic in Asia and Latin America.

In addition to transmission models like those described in this report, models of the immune system can play a significant role in our understanding of the AIDS epidemic. By adding to the understanding of the interactions of component parts of the immune system, these models can help guide vaccine and treatment developments. They may aid efforts to rid the infected cells of the virus so they can live a longer, healthier, happier life. One puzzle that models might help unravel is why T-4 cell counts are gradually depleted when apparently less than 0.1% are infected at any time and when the virus can stay dormant in such a cell for a long time. Depletion may be the result of quick destruction by cytotoxic T-cells or syncitia formation after infection, or it may be the result of excess HIV envelope proteins binding to CD4 sites and marking them for destruction. Macrophages, a significant fraction of which are infected in AIDS patients, probably provide a relatively indestructible reservoir of virus in either scenario [29]. Models may help to distinguish between several plausible scenarios or suggest experiments. Another puzzle is why chimpanzees, which can be infected with HIV but do not develop AIDS, seem to be able to fix and complete the complement sequence, while humans cannot [55]. Does complement effectively destroy the virus in chimps before the autoimmune response destroys the T-4 cells? Could our immune system be artificially stimulated to develop antibodies with a more effective complement procedure? Mathematical models can organize our understanding of the immune system in much the same way as the models described in this paper improved understanding of transmission.

Major advances are required before either an effective antiviral therapy or an effective vaccine is developed and becomes widely available. Thus, we have to prepare for a long battle against the spread of the AIDS epidemic. Our computer simulations of the transmission dynamics of the epidemic will

give us insight into how the epidemic is developing and will allow us to visualize the future.

The authors express their appreciation to W. A. Beyer, C. Castillo-Chavez, P. Denning, K. Dietz, J. S. Koopman, S. P. Layne, T. G. Marr, R. M. May, G. Myers, L. Sattenspiel, and B. K. Swartz for their comments and advice. Special thanks are also due to CDC for providing us with AIDS case data, C. R. Qualls for his thorough analysis of this data, Stirling Colgate for his valuable insights in understanding the cubic growth in AIDS cases, and Chriss Davis for her patience and skill in preparing this report. This research was supported by the Department of Energy under contracts W-7405-ENG-36, KC-07-01-01, and HA-02-02-02.

REFERENCES

1 J. Aron, Mathematical modeling of immunity to malaria, Proceedings of the 1987 CNLS Workshop on Nonlinearity in Medicine and Biology, *Math. Biosci.*, this issue.
2 D. Altman, *AIDS in the Mind of America*, Anchor Press/Doubleday, New York, 1986.
3 R. M. Anderson, R. M. May, G. F. Medley, and A. Johnson, A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS, *IMA J. Math. Appl. Med. Biol.* 3:229–263 (1986).
4 R. M. Anderson and R. M. May, Vaccination and herd immunity to infectious diseases, *Nature* 318:323–329 (1985).
5 R. M. Anderson and R. M. May, Directly transmitted infectious diseases: Control by vaccination, *Science* 215:1053–1060 (1982).
6 D. M. Auerbach, W. W. Darrow, H. W. Jaffe, and J. W. Curran, Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact, *Amer. J. Med.* 76:487–492 (1984).
7 J. L. Black, M. P. Dolan, H. A. DeFord, et al., Sharing of needles among users of intravenous drugs, Lancet *314*(I):467–447 (1986).
8 H. R. Brodt, E. B. Helm, A. Werner, et al., Spoutanverlauf der LAV/HTLV-III-Infektion, *Deutsche Med. Wochenschrift* 111:1175–1180 (1986).
9 D. S. Burke, J. F. Brundage, J. Herbold, et al., Human immunodeficiency virus (HIV) infections among civilian applicants for United States military service, in October 1985–March 1986: Demographic Factors Associated with Seropositivity (unpublished report), Nov. 1986.
10 Center for Disease Control, Results of Human T-lymphtropic virus type III test kits reported from blood collection centers—United States, April 22–May 19, 1985, *Morbidity and Mortality Weekly Report* 34(25):375–376 (1986).
11 S. A. Colgate, J. M. Hyman, and E. A. Stanley, A Risk Based Model Explaining the Cubic Growth in AIDS Cases, Los Alamos Nat. Lab. report, 1987.
12 W. W. Darrow, D. F. Echenberg, and H. W. Jaffe, Risk factors for human immunodeficiency virus infections in homosexual men, *Amer. J. Public Health* 77:479–483 (1987).
13 K. Dietz, The Dynamics of Spread of HIV Infection in the Heterosexual Population, unpublished report, 1987.

14 K. Dietz, On the transmission dynamics of HIV, Proceedings of the 1987 CNLS Workshop on Nonlinearity in Medicine and Biology, *Math. Biosci.*, this issue.
15 K. Dietz and D. Schenzle, Mathematical models for infectious disease statistics, in *A Celebration of Statistics* (A. C. Atkinson and S. E. Fienberg, Eds.), Springer, New York, 1985, pp. 167–204.
16 A. Finkbeiner, E. Hancock, and S. Schneider, AIDS, just the facts from specialists at Johns Hopkins, *Johns Hopkins Mag.*, Dec. 1986, pp. 15–27.
17 M. A. Fischl, G. M. Dickinson, G. B. Scott, et al., Evaluation of heterosexual partners, children, and household contacts of adults with AIDS, *J. Amer. Med. Assoc.* 257:640–644 (1987).
18 D. P. Francis, P. M. Feorino, J. R. Broderson, et al., Infection of chimpanzees with lymphadenopathy-associated virus, *Lancet*, (1 Dec. 1984), pp. 1276–1277.
19 H. M. Ginzburg, Intravenous drug users and the acquired immune deficiency syndrome, *Public Health Rep.* 99:206–212 (1984).
20 J. J. Goedert, S. H. Landesman, M. E. Eyster, and R. J. Biggar, AIDS incidence in pregnant women, their babies, homosexual men and hemophiliacs, presented at III International Conference on AIDS, Washington, 2 June 1987.
21 J. J. Goedert, M. Eyster, and R. Biggar, Heterosexual transmission of HIV: Association with severe T4-cell depletion in male hemophiliacs, presented at III International Conference on AIDS, Washington, 3 June 1987.
22 J. J. Goedert, R. J. Biggar, and D. M. Winn, Determinants of retrovirus (HTLV-III) antibody and immunodeficiency conditions in homosexual men, *Lancet*, 29 Sept. 1984, pp. 711–716.
23 R. Golubjatnikov, J. Pfister, and T. Tillotson, Homosexual promiscuity and the fear of AIDS, *Lancet*, 17 Sept. 1983, p. 681.
24 R. Grant, J. Wiley, and W. Winklestein, The infectivity of the human immunodeficiency virus: Estimates from a prospective study of a cohort of homosexual men, *J. Inf. Dis.* 156:189–193 (1987).
25 B. H. Hahn, M. A. Gonda, G. M. Shaw, et al., Genomic diversity of the acquired immune deficiency syndrome virus HTLV-III: Different viruses exhibit greatest divergence in their envelope genes, *Proc. Nat. Acad. Sci. U.S.A.* 82:4813–4817 (1985).
26 A. M. Hardy, E. T. Starcher, W. M. Morgan, et al., Review of death certificates to assess completeness of AIDS case reporting, *Public Health Rep.* 102:386–391 (1987).
27 N. A. Hessol, G. W. Rutherford, P. M. O'Malley, et al., The natural history of human immunodeficiency virus infection in a cohort of homosexual and bisexual men: A 7-year prospective study, presented at III International Conference on AIDS, Washington, 1 June 1987.
28 H. W. Hethcote and J. A. Yorke, Gonorrhea: Transmission dynamics and control, *Lecture Notes Biomath.* 56:1–105 (1984).
29 D. D. Ho, R. J. Pomerantz, and J. C. Kaplan, Pathogenesis of infection with human immunodeficiency virus, *New Engl. J. Med.* 317:278–286 (1987).
30 J. M. Hyman, C. R. Qualls, and E. A. Stanley, Analysis of CDC AIDS Data, Los Alamos Nat. Lab. report, in preparation.
31 J. Jacquez, J. Koopman, L. Sattenspiel, C. Simon, and T. Perry, Modeling and the transmission of HIV, in preparation.
32 J. G. Joseph, S. Montgomery, R. C. Kessler, et al., Two-year longitudinal study of behavioral risk reduction in a cohort of homosexual men, presented at III International Conference on AIDS, Washington, 1 June 1987.

33 F. N. Judson, Fear of AIDS and gonorrhea rates in homosexual men, *Lancet*, 16 July 1983, pp. 159–160.
34 L. A. Kingsley, R. Kaslow, C. R. Rinaldo, Jr., et al., Risk factors for seroconversion to human immunodeficiency virus among male homosexuals, *Lancet* 21 Feb. 1987, pp. 345–349.
35 I. Kolb, *Das Kreuz mit der Liebe*, Gruner and Jahr, Hamburg, 1980.
36 J. M. A. Lange, D. A. Paul, H. G. Huisman, et al., Persistent HIV antigenemia and decline of HIV core antibodies associated with transition to AIDS, *British Med. J.* 293:1459–1462 (1986).
37 K. Lui, D. N. Lawrence, W. M. Morgan, et al., A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome, *Proc. Nat. Acad. Sci. U.S.A.* 83:3051–3055 (1986).
38 J. L. Martin, The impact of AIDS on gay male sexual behavior patterns in New York City, *Amer. J. Public Health* 77:578–581 (1987).
39 E. Martini, Betrachtungen zur Epidemiologie der Malaria und der Syphilis, *Dermatol. Wochenschrift* 19:640–643 (1928).
40 R. M. May and R. M. Anderson, Transmission dynamics of HIV infection, *Nature* 326:137–142 (1987).
41 R. M. May, R. M. Anderson, and A. R. McLean, Possible demographic consequences of HIV/AIDS epidemics: I. Assuming HIV infection always leads to AIDS, Proceedings of the 1987 CNLS Workshop on Nonlinearity in Medicine and Biology, *Math. Biosci.*, this issue.
42 G. F. Medley, R. M. Anderson, D. R. Cox, et al., Incubation period of AIDS in patients infected via blood transfusion, *Nature* 238:719–721 (1987).
43 M. Melbye, R. J. Biggar, P. Ebbesen, et al., Long-term seropositivity for human T-lymphotropic virus type III in homosexual men without the acquired immunodeficiency syndrome: Development of immunologic and clinical abnormalities, *Ann. Int. Med.* 104:496–500 (1986).
44 J. D. Murray, E. A. Stanley, and D. L. Brown, On the spatial spread of rabies among foxes, *Proc. Roy. Soc. London Ser. B* 229:111–150 (1986).
45 N. Padian, J. Wiley, and W. Winkelstein, Male to female transmission of human immunodeficiency virus: Current results, infectivity rates, and San Francisco population seroprevalence estimates, presented at III International Conference on AIDS, Washington, 4 June 1987.
46 T. A. Peterman, R. L. Stoneburner, J. R. Allen, et al., Risk of human immunodeficiency virus transmission from heterosexual adults with transfusion-associated infections, *J. Amer. Med. Assoc.* 259:55–58 (1988).
47 R. Redfield, The clinical research and public health implications of the Walter Reed staging classification of HIV infection, presented at III International Conference on AIDS, Washington, 1 June 1987.
48 R. R. Redfield and D. S. Burke, Shadow on the Land: The Epidemiology of HIV Infection, unpublished report, 1987.
49 M. F. Rogers and W. W. Williams, AIDS in Blacks and Hispanics: Implications for prevention, *Issues Sci. and Technol.*, Spring 1987, pp. 89–94.
50 R. Ross, *The Prevention of Malaria*, 2nd ed., Murray, London, 1911.
51 L. Sattenspiel, Population structure and the spread of disease, *Human Biol.* 59:411–438 (1987).
52 L. Sattenspiel and C. Simon, The spread and persistence of infectious diseases in structured populations, Proceedings of the 1987 CNLS Workshop on Nonlinearity in Medicine and Biology, *Math. Biosci.*, this issue.

53 T. Smith, M. Marcus, and G. Myers, Phylogenetic analysis of HIV-1 and HIV-2, presented at Cold Spring Harbor Conference on AIDS, New York, Sept. 1987.

54 S. Z. Sulahuddin, P. D. Markham, R. R. Redfield, et al., HTLV-III in symptom-free seronegative persons, *Lancet*, 22–29 Dec. 1984, pp. 1418–1420.

55 S. H. Weiss, P. R. Clapham, J. N. Weber, et al., Variable and conserved neutralization antigens of human immunodeficiency virus, *Nature* 324:572–575 (1986).

56 S. H. Weiss, W. C. Saxinger, D. Rechtman, et al., HTLV-III infection among health care workers. Association with needle-stick injuries, *J. Amer. Med. Assoc.* 254:2089–2093 (1985).

57 J. A. Wiley, G. W. Rutherford, A. R. Moss, et al., Age and cumulative evidence of AIDS among seropositive homosexual men in high incidence areas of San Francisco, presented at III International Conference on AIDS, Washington, 5 June 1987.

58 W. Winklestein, M. Samuel, N. S. Padian, et al., The San Francisco Men's Health Study: III. Reduction in HIV transmission among homosexual/bisexual men 1982–86, *Amer. J. Public Health* 76:685–689 (1986).

59 W. Winklestein, D. M. Lyman, N. S. Padian, et al., Sexual practices and risk of infection by the human immunodeficiency virus, *J. Amer. Med. Assoc.* 257:321–325 (1987).