

# Modeling the AIDS Epidemic

## *Mathematical Techniques Provide Qualitative Insights*

A team of mathematicians at Los Alamos National Laboratory has been using sophisticated mathematical techniques and high-powered computing to formulate a set of models of the AIDS epidemic. Not only are these models more consistent with existing data than previous models, they also provide some advantages over statistical methods, which are based only on past trends. The qualitative insights the models provide allow researchers to test various hypotheses about the disease and pinpoint effective strategies to stem the spread of AIDS. In addition, the models can be used to unify existing data and guide the collection of new data.

### **Cubic Growth of AIDS**

One of the most surprising aspects of the AIDS epidemic is that, unlike most epidemics, it does not exhibit exponential growth. Rather, the best fit to the data on the number of AIDS cases to date turns out to be a cubic polynomial. This phenomenon was seen in the first models of the AIDS epidemic, formulated by the Centers for Disease Control (CDC) in 1986. "I don't know of any other epidemic that's growing polynomially," says James M. Hyman, one of the Los Alamos modelers. "Usually it's a very quick exponential, saturates out, and dies back down." Even when the population is split into groups according to race, age, gender, or even region of the country, the number of AIDS cases still exhibits this cubic polynomial behavior (Figure 1). "Cliff Qualls [a statistician at the University of New Mexico] noticed that no matter how you cut the data, it always had this cubic growth," remarks Hyman. "Whatever the mechanism that's keeping it from being exponential, it's probably similar in the different categories."

Why does AIDS differ in this way? The main reason appears to be the social factors that play a large role in the spread of the human immunodeficiency virus (HIV, the virus causing AIDS). For a disease like malaria, transmission is randomly distributed over the population, essentially because mosquitos don't care who they bite. However, the behaviors that put people at risk of

exposure to HIV are not randomly distributed, but tend to be confined to certain segments of the population.

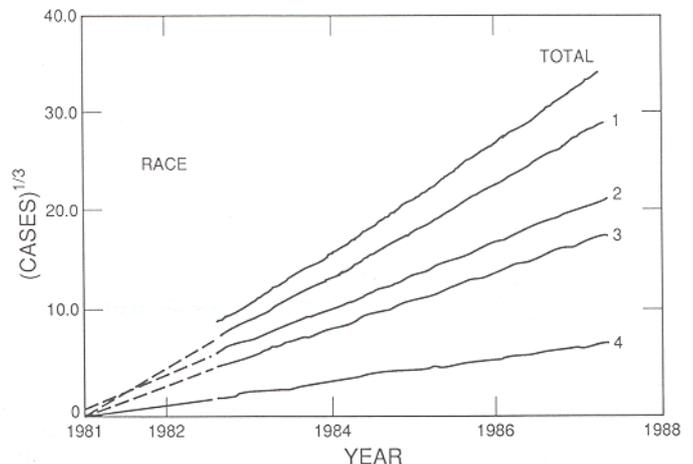


Figure 1. Data reported by the CDC. Shown here is the cube root of the cumulative number of AIDS cases plotted versus time. Also shown are the same data broken down by racial subgroups: 1. White, 2. Black, 3. Hispanic, and 4. Unknown.

### **Biased Mixing Model**

The primary mechanisms for transmission of HIV are sexual contact and sharing of intravenous needles. Since little is known about needle-sharing habits, the Los Alamos group has focused on modeling sexual contact. The model divides the population into risk categories designated by different behaviors correlated with HIV infection, such as sexual partner exchange rate. They also adopted the assumption of "biased mixing"—the idea that people do not choose sexual partners at random, but tend to choose from others in their own risk group. For example, the biased-mixing model assumes that those with a high partner-exchange rate tend to mix with others having a large number of partners, while those who tend to be monogamous mix with others who also have few sexual partners. Within each risk group, mixing is assumed to be random.

With these assumptions, “you get a completely different epidemic” than with random mixing, notes Hyman. The model predicts that people in the high-risk groups become infected first, and they tend to infect others in their group. The peak of the number of infected people occurs first in the high-risk groups, and moves in a “saturation wave” to lower-risk groups (Figure 2). The model is consistent with the data, for it predicts the number of AIDS cases will increase like a cubic polynomial. “What happens is that the social behavior, once we put it into the model, slows the epidemic down tremendously” from the usual exponential growth, says Hyman.

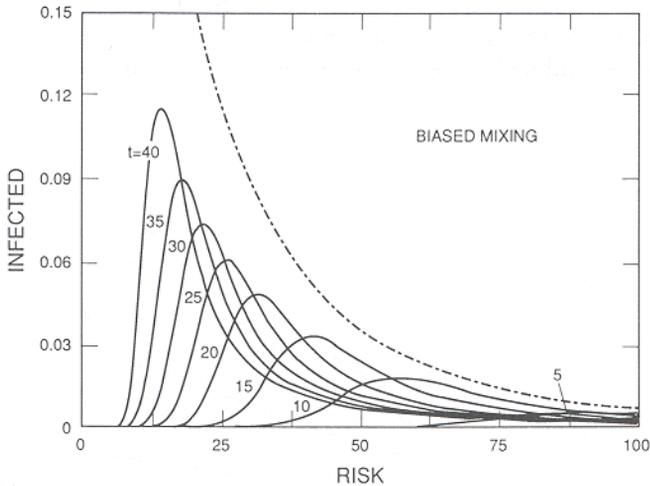


Figure 2. Numerical results of the biased mixing model, showing the fraction of the population infected versus risk, defined here as the number of sexual partners per year. Each solid curve represents the distribution of infection at a particular time, with  $t$  measured in years. The dotted line shows the distribution of risk in the population.

In contrast, if one assumes random mixing among people in high- and low-risk groups, then a very different picture emerges. The epidemic actually spreads much more rapidly than if one assumes biased mixing. This is because the majority of the population is concentrated in the lower-risk groups, and, with random mixing, many more of these people are exposed to the virus (Figure 3). This model is inconsistent with current thinking about the epidemic, because it shows the peak in the number of infected beginning and staying within the lower-risk groups. In addition, this model predicts an exponential, rather than cubic, growth in the number of AIDS cases.

Because of these results, the modelers knew that social structure was crucial to their models. In addition to the number of partners per year, important parameters include age, sexual contact frequency, and the probability of infection. The Los Alamos team began with a logistic equation describing the change over time in the number of infected,  $I(t)$ ,  $dI/dt = \alpha I(1 - I/N)$ . Here,  $\alpha$  is the rate at which an infected person passes on the virus and  $N$  is the size of the population. This is the basic equation of epidemiology describing the spread of a virus based

on random infection. The team built their model by successively modifying this equation to accommodate the various parameters.

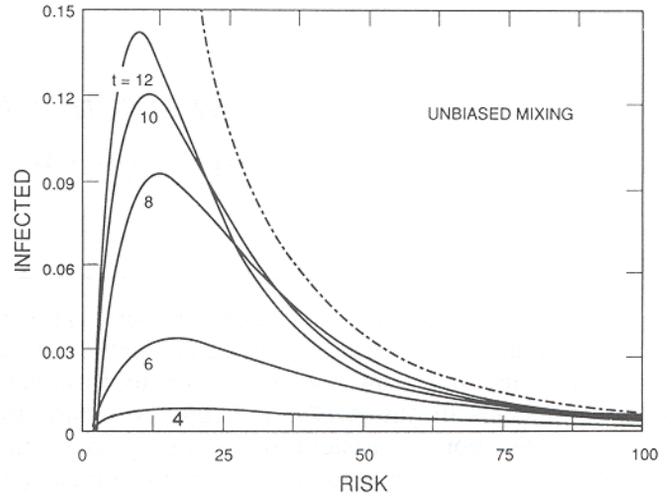


Figure 3. Same as Figure 2, but with the assumption of random mixing.

One of the parameters the model is most sensitive to is transmissibility, a quantity, related to  $\alpha$ , that measures the likelihood that an infected person will transmit the virus to another individual. Although the way transmissibility varies over time is not well understood, there is evidence to suggest that HIV-carriers are most infectious immediately after they have been infected. After that, transmissibility drops off before rising again as the virus begins to destroy the individual’s immune system. The time since infection strongly influences the way the virus spreads and therefore plays an important role in modeling. The modelers have experimented with using functions with sharp initial peaks, constant functions, and linear functions for the transmissibility of HIV. All of these give strikingly different pictures of the distribution of infection in the population. The sensitivity of the model to transmissibility indicates the crucial need for more data on the probability of transmission of HIV.

Data on sexual behavior and mixing patterns among and within various groups is also sorely lacking. In fact, the modelers have had to rely on the 1948 Kinsey report for data on various aspects of sexual behavior, such as sexual outlet frequency, because more current data simply are not available. Because of privacy concerns, gathering data on sexual behavior is rarely a simple task. In the socially and politically charged atmosphere surrounding the AIDS epidemic, such concerns are even more pronounced.

### Computing Power Imposes Limits

Although the Los Alamos model can allow for a great deal of detail, current computing power imposes limitations.

For a model with  $d$  parameters that typically range over 100 different values, the number of data points is  $100^d$ . Standard numerical techniques work well for solving the differential equations in the model, but if  $d$  is larger than 3 or 4, the size of the problem jumps out of the range of current computing power. "You very quickly see that when we add four more parameters [for a total of 7 or 8], we run into trouble," says Hyman. "There's no way to hit it with brute force. I'm using a Cray Y-MP and I can't even touch it. Three-dimensional fluid dynamics looks easy" compared to this. "What's needed are analytical tools to reduce the complexity." Hyman says he and his Los Alamos colleague, E. Ann Stanley, are investigating the use of a separation of variables technique that would allow some of the degrees of freedom to be collapsed, thereby reducing the computational load. They hope the technique will allow problems now requiring the Cray to be run on a Sun workstation.

In addition to the Los Alamos team, several other mathematicians are working on AIDS epidemiology. The annual conference of the Society for Industrial and Applied Mathematics, held in San Diego this past July, featured a minisymposium on the mathematical modeling of the AIDS epidemic. In addition to Stanley, who spoke on the work at Los Alamos, Herbert Hethcote of the University of Utah described his research, which applies dynamic simulation methods to early data on HIV infection among gay men in San Francisco. Carl Simon of the University of Michigan spoke on sophisticated methods for modeling nonrandom mixing, and Carlos Castillo-Chavez of Cornell University talked about aggregation and mixing in HIV transmission. Apart from the epidemiology of AIDS, mathematicians are also working on modeling other aspects of the disease, such as the way HIV mutates, its 3-dimensional structure, and the way it destroys the immune system.

The Los Alamos group is involved in a project to develop user-friendly software that will allow public health workers to use the model to better understand the future of the epidemic. For this project, researchers at the University of Illinois are working on the software, the Census Bureau is providing data, and the Air Force Academy will be testing the package. The package will eventually be used as part of a demographic model of Africa, in which the AIDS epidemic will be one component.

### Qualitative Insights

Hyman notes that the usefulness of the models lies "not so much in quantitative projections, but qualitative insights for questions like, is it better for us to spend our

money on condoms, or should we go after other [sexually transmitted diseases]? Are clean needles more important than a clean blood supply?" The models can be helpful in predicting which strategies will be the most effective.

Although he sees great promise in modeling, Hyman is cautious about overselling it. "Our biggest fear is that people will take the numbers too absolutely," he says. "The main strength of models lies in predicting qualitative behavior." Statistical extrapolation is still the best way to predict the number of AIDS cases over the next few years. The Los Alamos model has been very accurate, with predictions that have matched the data within 2-3%, but the models' main advantage is that they allow researchers to test various hypotheses. "With extrapolation, you can't address such questions as, what if we double the number of people who use condoms, how would that change the predictions?" Hyman remarks. "There's no way to put that into your least-squares fit." In addition, statistical extrapolations are probably reliable only for predictions over 5 years, while the models are capable of predicting qualitative trends over much longer periods.

Some researchers at the CDC and others in the medical sciences community were at first skeptical of the usefulness and validity of the mathematical models. "Good modelers in the field are really looking for qualitative understanding," says Hyman. "Models at this point can't do better than extrapolation at guessing the number of AIDS cases, which is what CDC in the past had been most interested in." But now, he says, attitudes are changing. "The whole point of the model is to get people to think in terms of the future, not just in terms of how many AIDS cases we have now, but in terms of how many infected people we have now, and what that's going to mean further down the road ... People are dying, people are getting infected every day. With estimates of over one million people currently infected, if mathematical models can give any new insights, it could result in a tremendous saving of lives."

#### Further reading:

"AIDS and a risk-based model," by Stirling A. Colgate, E. Ann Stanley, James M. Hyman, Clifford R. Qualls, and Scott P. Layne, *Los Alamos Science*, Fall 1989.

"Los Alamos Researchers Model AIDS Epidemic," by Stirling A. Colgate, James M. Hyman, and E. Ann Stanley, *SIAM News*, May 1989.

"The Science of Computing: Modeling the AIDS Epidemic," by Peter J. Denning, *American Scientist*, November/December, 1988.

**Allyn Jackson**  
Staff Writer