

NBER WORKING PAPER SERIES

PROJECTING THE NUMBER OF NEW  
AIDS CASES IN THE U.S.

David E. Bloom

Sherry Glied

Working Paper No. 4180

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 1992

The authors thank Bowen Garrett and Cynthia Miller for helpful computational assistance. The authors also thank two anonymous referees for their comments as well as seminar participants at the HIV Center and at the Econometrics Workshop at Columbia University and at the 1992 Summer Institute in Health Economics at the National Bureau of Economic Research. This research was supported by NIH Grant HD25914-01. This paper is part of NBER's research program in Health Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

PROJECTING THE NUMBER OF NEW  
AIDS CASES IN THE U.S.

ABSTRACT

This paper reviews the two leading methods used to project the number of AIDS cases: back calculation and extrapolation. These methods are assessed in light of key features of the HIV/AIDS epidemic and of data on the epidemic; they are also assessed in terms of the quality of the projections they yield.

Our analysis shows that both methods have tended to overproject, often by sizable amounts, the number of AIDS cases in the U.S., especially among homosexual/bisexual males and users of blood and blood products. Our results provide no evidence that the use of AZT and other prophylaxis accounts for these projection errors. Rather, the overprojections appear to be mainly the result of a considerable reduction in the rate of new HIV infection among the gay community starting in 1983-85.

A new method for projecting AIDS cases is proposed that exploits knowledge about the process generating AIDS cases and that incorporates readily available information about rates of new HIV infection. This method is far less sensitive to estimates of the incubation distribution than the method of back calculation and is shown, for the two transmission categories studied, to generate far more accurate AIDS case projections through 1990 than those based on the method of extrapolation. Relative to the method of extrapolation, this method projects 22,000 fewer new AIDS cases for 1995 (a 36 percent difference). This method also projects that intravenous drug users will replace homosexual/bisexual men as the dominant transmission category for AIDS.

David E. Bloom  
Department of Economics  
Columbia University  
New York, NY 10027  
(212) 854-3680  
and NBER

Sherry Glied  
School of Public Health  
and  
Department of Economics  
Columbia University  
New York, NY 10027  
(212) 854-3926  
and NBER

## Projecting the Number of New AIDS Cases in the U.S.

Between 1981 and 1989 the reported number of deaths from AIDS in the U.S. increased by roughly two orders of magnitude. In total, over 80,000 Americans are reported to have died of AIDS during the 1980s (Centers for Disease Control, 1992). Although this total represents less than 1/2 of 1 percent of all deaths in the U.S. during that decade, the sharp rate of increase has provoked great fears of massive numbers of future deaths due to AIDS. Indeed, early projections from the Centers for Disease Control (CDC) indicated that cumulative totals of 270,000 AIDS cases and 180,000 AIDS deaths would occur by the end of 1991 (Public Health Service, 1986).

In combination with the actual rise in AIDS mortality, the large numbers of projected AIDS cases and AIDS deaths have led to major responses among individuals, employers, hospitals, insurance companies, and public policymakers. Rates of sexual behavior that increase the risk of HIV infection appear to have declined, at least among certain groups (see Becker and Joseph, 1988). Many hospital systems have expanded their capacity to deal with HIV/AIDS on both an inpatient and an outpatient basis (see Hendrix, 1987 and Volberding, 1987). Insurance companies have established coverage limits and now typically require negative HIV tests as a precondition for certain types and amounts of individual coverage (see Eden, 1988 and MetLife, 1990). Public policymakers have increased funding for biomedical research related to the detection, prevention, and treatment of HIV infection, mounted campaigns to educate the public about HIV/AIDS, and implemented legislation to protect individuals from AIDS-related discrimination in housing and labor markets (see Rhein, 1990 and Rayhawk, 1991). Policymakers have also devoted considerable resources to the care of HIV-infected individuals.

Given the wide range of costly social, legal, and economic adjustments that have been made at least partly on the basis of widely-publicized AIDS case projections, a critical review of the methods used to calculate those projections seems well in order. Section I describes the two most-widely used techniques for projecting AIDS cases: extrapolation and back calculation. Section II describes key data imperfections that may affect the quality of the projections these techniques yield. These imperfections include delays in the reporting of AIDS cases to the CDC, changes in the official criteria used to define an AIDS case, and incomplete reporting to the CDC of individuals who have been diagnosed as having AIDS. Refinements of the two basic projection methods to account for these data imperfections, which has been the main focus of the statistical literature in this area, are also discussed in this section. Section III provides an empirical assessment of the two techniques by comparing previously-constructed projections of the incidence of AIDS (for the entire U.S. population) to each other as well as to actual data on the incidence of AIDS.

We find that most researchers have over-projected, often by sizable amounts, the number of AIDS cases subsequently recorded in the U.S., particularly for certain transmission categories. Recent studies have argued that advances in the medical treatment of HIV/AIDS can potentially explain these projection errors (see Gail, Rosenberg, and Goedert, 1990). We test this view and find little evidence to support it. Instead, our analysis suggests that the overprojections are largely due to reductions in the rate of new infections in the early to

mid-1980s not accounted for in existing projections. In Section IV we propose a new method for exploiting data on the rate of new infections and we implement this method using data on two cohorts. The simple method we propose clearly outperforms most other projections based on the techniques of back calculation and extrapolation. Section V summarizes our results and offers some suggestions for further research.

## I. Methods of Projecting the Incidence of AIDS

Reported AIDS cases are the culmination of a series of events that begin with specific individual behaviors including needle sharing among intravenous drug users, engaging in heterosexual or homosexual sex, and receiving a blood transfusion. Infection by the human immunodeficiency virus (HIV) is a probabilistic outcome of these behaviors. The presence of the HIV (or, more properly, antibodies to the HIV) can be detected in an individual's bloodstream as early as four to six weeks after infection and will, if untreated, lead to symptoms of AIDS as early as two years after infection. The onset of particular symptoms, in combination with a positive test result, is required for a physician to diagnose an individual as having AIDS and to report that diagnosis to the state or local public health authority, which is responsible for reporting it to the CDC.

One method of projecting AIDS cases, known as micro-simulation, attempts to characterize statistically each stage of this process. Although projections of AIDS cases on the basis of such a method have considerable intuitive appeal, the utility of this technique is, as a practical matter, quite limited because many key parameters necessary for its implementation are highly uncertain, such as the size of the different sub-populations affected by the epidemic and rates of interaction within and between those sub-populations (see Gail and Brookmeyer, 1988 and United States General Accounting Office, 1989). [1] More commonly, projections of the incidence of AIDS are based on statistical methods that use a more limited information set. Two such methods are now widely used: extrapolation and back calculation. Both are described below (see Gail and Brookmeyer, 1988; Brookmeyer, 1991; United States General Accounting Office, 1989; and Hellinger, 1990 for more detailed descriptions).

### (a) Extrapolation

The simplest version of projection by extrapolation involves regressing a time series of AIDS cases on a trend variable and using the estimated parameters to generate projections of future cases. More complicated versions of this method employ specifications that include nonlinear time trends or that use nonlinear functions of the number of AIDS cases as the dependent variable. Separate vectors of coefficients can be estimated for different subgroups of the population (e.g., homosexual/bisexual men (HBM), intravenous drug users (IVDU), etc.) and used to construct group-specific projections of the incidence of AIDS that can then be combined to form refined aggregate projections.

The main advantages of extrapolation are its simplicity, its flexibility, and the fact that it can be used to project the incidence of AIDS as far into the future as one is willing to assume a continuing stable process. The main disadvantage relates to this method's mechanical nature, which provides little natural encouragement to think about and incorporate information directly related to the underlying behavioral and epidemiological processes generating AIDS cases.

(b) Back calculation

The method of projecting AIDS cases by back calculation proceeds in two steps. First, a time series of new AIDS cases is mapped backwards into the number of HIV infections that must have occurred in prior periods to generate that time series. This "back calculation" is performed using independent information on the distribution of time between infection with the HIV and the emergence of AIDS symptoms. The most popular sources of information on this "incubation distribution" are relatively small-sized samples of individuals in particular transmission categories, with many cases that had yet to progress from HIV infection to full-blown AIDS at the time that they were observed (see Gail and Brookmeyer, 1988; United States General Accounting Office, 1989; Brookmeyer and Liao, 1990; and Centers for Disease Control, 1990).

Second, the estimated time series of HIV infections (that is back calculated from the time series of AIDS cases) is forward projected using the incubation distribution to reconstruct the observed time series of AIDS cases and to generate projections for future time periods. Projections constructed in this manner will underestimate the number of future AIDS cases if new infections occurred beyond the period for which infection rates can be inferred based on the time series of AIDS cases. However, the fact that the proportion of cases in which there is less than a two-year lag between HIV infection and symptomatic AIDS is close to zero limits the extent to which new infections introduce error into one- and two-year projections. For constructing projections further than two years into the future, one could rely upon independent estimates of the number of new infections. However, in practice, the number of new infections is usually assumed to be zero, undermining the accuracy of the resulting projections beyond two years into the future, potentially by large amounts. On the other hand, and in contrast to the method of extrapolation, the method of back calculation makes no assumptions at all about the stability of the infection-generating process. (Table 1 illustrates the method of back calculation with a simple hypothetical example.)

The accuracy of projections produced by the method of back calculation is critically dependent on the quality and stability of data on the distribution of delay between HIV infection and the onset of AIDS. Unfortunately, existing estimates of the incubation distribution may be inaccurate because (1) they are based on censored samples drawn years ago from a population distribution of unknown form, (2) they may vary across transmission categories that have themselves changed in relative importance over time, and (3) incubation distributions for given transmission categories may be unstable over time, especially since the advent of new drug therapies for the treatment of HIV infection. Because AIDS projections are highly sensitive to the assumed incubation distribution, these potential sources of error are serious weaknesses of

the method of back calculation. Indeed, most of the widely-used incubation distributions produce estimates of negative numbers of infections in one or more years, a patently absurd result that has led to the imposition of nonnegativity constraints in recent applications of this method (see Hay and Wolak, 1990 and Hellinger, 1990).

Projections based on the method of back calculation are particularly sensitive to the incubation distribution because that distribution is used twice in constructing the projections: first to back calculate the number of infections and then to forward calculate the number of AIDS cases from the time series of estimated infections. To illustrate this sensitivity, we perturb slightly the hypothetical incubation distribution used in Table 1 and construct a new projection of the number of AIDS cases for time period 9. The perturbed distribution is as follows: 20 percent of individuals develop AIDS symptoms three years after they are infected, 40 percent four years after infection, and 40 percent after five years. Using this perturbed distribution yields a projection of 3700 cases for time period 9 (a 40 percent (average) deviation between this projection and that based on the original incubation distribution).

## II. Data and Methodological Issues

The preceding section described the two leading methods used to project the number of AIDS cases. The discussion assumed that the data used to implement these models are accurate measures of the number of AIDS cases diagnosed in each period. Although CDC reporting of AIDS cases is remarkably complete and timely, the AIDS case data are nonetheless imperfect in a number of ways that may have important effects on the quality of the projections they inform.

To illustrate these problems, consider the following hypothetical system for the surveillance of AIDS. On the day that every HIV-infected individual first manifests symptoms of AIDS, he or she sees a physician who correctly diagnoses his or her illness and reports the illness directly to the local or state public health department, which immediately reports this information to the CDC. Assume further that physicians make no incorrect diagnoses of AIDS and that the criteria for diagnosing AIDS are unchanging over time.

Contrast this (almost ideal) AIDS surveillance system with the one actually in place in the U.S. First, under the actual system there is a variable lag between the onset of AIDS symptoms and the diagnosis of AIDS. Indeed, some individuals suffering from AIDS may never be officially diagnosed as such, perhaps because (1) they die from some opportunistic infection or other cause before an AIDS diagnosis is made, (2) they and their physician suppress the true diagnosis in an effort to avoid any stigma associated with being identified as an AIDS sufferer, or (3) their physician fails to make a correct diagnosis.

Second, because incentives to comply promptly with AIDS reporting requirements are weak, physicians (or the institutions with which they are affiliated) do not, in practice,

immediately report all AIDS diagnoses to state public health departments and state public health departments do not immediately report all AIDS diagnoses to the CDC. The time lag between the diagnosis and reporting of an AIDS case is variable and may be lengthy, with some cases perhaps never being reported.

Third, the criteria that must be met to justify an official AIDS diagnosis have changed over time, with the number of symptoms generally being expanded and the need for laboratory confirmation of AIDS symptoms being eliminated (see United States General Accounting Office, 1989).

Under these circumstances, the CDC's time series of newly-diagnosed AIDS cases will (1) understate the true number of AIDS cases (because some cases are never diagnosed and some diagnoses are never reported); (2) understate the true number of AIDS diagnoses (because some diagnoses are never reported while others are reported with a delay); and (3) reflect inconsistent information over time (if the delay distributions between the onset of symptoms, the physician's diagnosis, and the CDC case report change over time or if the definition of AIDS changes).

These features of the CDC's AIDS data must be considered when using those data to project the national incidence of new AIDS cases. By comparing death certificate records to CDC case reports, the extent of underreporting has been estimated to be roughly 15 percent (with substantial regional variation, from 10 percent in major cities to 40 percent in low-prevalence areas; see Hardy, et al. 1987; Conway, et al. 1989; and Centers for Disease Control, 1990). Ideally, reported AIDS cases can be inflated by this underreporting rate to generate estimates of total cases. However, the underreporting rate may be unstable over time, especially as physicians become more familiar with HIV/AIDS, less stigma is attached to AIDS, Medicaid reimbursement rates for AIDS cases are increased, states become more aggressive in their surveillance of AIDS, and the geographic distribution of the disease shifts. Note also that the true number of AIDS cases may also be understated by death certificate data.

Reporting delays are known to be lengthy in the case of AIDS. For example, 2.3 percent of cases diagnosed in 1982 and reported to the CDC through 1985 were reported in 1985. Available data on the distribution of time between the rendering of AIDS diagnoses and the receipt of AIDS case reports at the CDC can be used to inflate cases reported into cases diagnosed. Unfortunately, the distribution of reporting lags has been shown to be unstable over time, leaving the accuracy of these adjustments in some doubt (Harris, 1990). To illustrate, 4.3 percent of cases diagnosed in 1987 and reported to the CDC through 1990 were reported in 1990 (an 85 percent increase from the 1982-85 example cited above). [2] Furthermore, the delay between the onset of AIDS symptoms and the rendering of an AIDS diagnosis is also unknown. Given the existence of life-prolonging drug therapies for the treatment of AIDS, this delay is also likely to have shortened over time, especially since reimbursement for the expense of those drugs often requires a definitive AIDS diagnosis.

The CDC's adoption of an expanded AIDS case definition in 1987 led to an estimated 18 percent increase in the number of reported AIDS cases (see Gail and Brookmeyer, 1988 and

Centers for Disease Control, 1990). In order to use pre- and post-1987 data for projection, it is necessary to adjust them for consistency before and after this definitional change, a complex task whose handling has been, at best, imperfect (see Karon, Dondero, and Curran, 1988 and Karon, Devine, and Morgan, 1989). Given the difficulties of adjusting the data, some projections are based just on data recorded either before or after the definitional change (see Gail and Brookmeyer, 1988 and Hellinger, 1988). [3]

The data problems described above all represent potential sources of projection error. Underreporting will tend to lead to underprojections of the true number of actual AIDS cases using both the methods of extrapolation and back calculation but to accurate projections of reported AIDS cases unless the extent of underreporting changes over time. For example, if the extent of underreporting diminishes over time, projections based on either extrapolation or back calculation are likely to overpredict the true number of reported AIDS cases.

Reporting lags can introduce a wide range of errors into AIDS projections whether those projections are constructed via extrapolation or back calculation. However, the nature and magnitude of the errors will depend, in general, upon the specific pattern of reporting lags and its stability over time.

Changes in the criteria used to define an AIDS case have complex and uncertain effects on AIDS projections. Definitional changes are akin to a reduction in underreporting in terms of their effect on the total number of cases ever reported; they are also akin to a decline in reporting delay in terms of their effect on the time pattern of diagnosed cases.

### III. Review and Assessment of Earlier Projections

Table 2 summarizes the leading studies that report AIDS case projections for 1991. The studies fall into two groups according to whether the projections are based on the method of back calculation or extrapolation. All of the projections reported in Table 2 refer to cases that would be diagnosed in 1991 (and that would eventually be reported to the CDC). In addition, all of the studies base their projections on the CDC's monthly HIV/AIDS surveillance data. Although most of the studies inflate their projections to account for the underreporting of AIDS cases, those adjustments have been undone in the preparation of Table 2 in order to report figures that are comparable to the number of AIDS cases that will ultimately be reported as having been diagnosed in 1991. All of the projections are based on data that have been adjusted for reporting delays, although the methods used to make those adjustments vary across the studies. Finally, four of the five studies in Table 2 are based on the pre-1987 AIDS case definition.

The third and fourth columns of Table 2 report projections from each study of the number of AIDS cases that would be diagnosed in 1991. The third column reports each study's actual projection for 1991, after removing the effect of any adjustment for underreporting made in the

study. The fourth column adjusts these figures so that they represent projections of the number of diagnosed AIDS cases in 1991 based on the post-1987 case definition (i.e., the adjustment is made by multiplying the projections in column three by 1.18, following Gail and Brookmeyer, 1988).

The projections reported in Table 2 range from 53,000 to 89,000 cases (adjusted for the 1987 case definition change). This range exceeds by 15 to 93 percent our estimate of the actual number of cases that were diagnosed in 1991 and will ever be reported (i.e., 46,000). [4, 5, 6]

Figure 1 plots monthly AIDS cases from 1982 to 1990, along with a projection curve from 1987 to the end of 1990. The data plotted have been adjusted by the CDC for reporting delay (as of June, 1991). The projection curve is derived by regressing these data from 1982 through the end of 1986 on a quadratic function of time (i.e., a standard extrapolation model). Figure 2 presents a similar plot using quarterly AIDS case data that have been adjusted by the CDC for the 1987 change in the AIDS case definition. (These plots of "definitionally consistent cases" cover the period from the first quarter of 1983 through the third quarter of 1990, with the regression line estimated using data through the fourth quarter of 1986.) [7]

Figures 1 and 2 illustrate graphically that the extrapolation method leads to overprojections of the number of AIDS cases, with the divergence becoming quite substantial by 1990. Although not shown in these figures, extrapolation models based on more elaborate specifications (e.g., such as those reported in Hellinger 1988; United States General Accounting Office, 1989; and Hellinger, 1990) show a similar pattern.

To gain further insight into the origin of the projection errors revealed in Figures 1 and 2, we examine two studies that report AIDS case projections for specific transmission categories. Brookmeyer and Damiano, 1989 (which bases its projections on back calculation, assuming no new infections after 1987) and Morgan and Curran, 1986 (which bases its projections on extrapolation), both overpredict the share of AIDS cases involving HBMs and underpredict the share of AIDS cases involving IVDUs that would occur in 1991. [8] Morgan and Curran also report projections for specific regions in 1991. They overproject the proportion of AIDS cases occurring in San Francisco, where HBM is the dominant transmission category, and underproject the share of cases in New York City and Florida, where the dominant transmission category is IVDU.

Figures 3a, 3b, and 3c plot monthly AIDS cases from 1982 through 1990 for three transmission categories: HBMs, blood transfusion and blood products cases (BTBP), and IVDUs. [9] Figures 4a, 4b, and 4c plot quarterly data on definitionally consistent AIDS cases. The plots in Figures 3 and 4 reveal quite clearly that the flattening of the time series of all AIDS cases was associated most closely with a leveling off in the number of cases involving HBM and BTBP that appears to have begun in mid-1987. [10] Indeed, quadratic projection curves fit separately to the monthly and quarterly data for the three transmission categories indicate that the largest source of overall error is due to HBM. Tests for the stability of these time-series regressions before and after the beginning of 1987 are reported in Table 3, using both the monthly and quarterly data.

All of the tests for the HBM, BTBP, and IVDU cases lead to rejections of the null hypothesis of stability. [11] These findings undermine the confidence that can be placed in the extrapolation method because they suggest that the fundamental premise of stability upon which this method is based is violated. [12]

#### IV. Explaining the Forecast Errors

The preceding section demonstrated that extrapolation models, widely used for constructing AIDS case projections (1) have unstable parameters, (2) have led to sizable overprojections of the number of AIDS cases, and (3) have provided especially poor projections for HBM and BTBP cases, as early as 1987. We also showed that back calculation models, which are inherently limited in their ability to project more than two or three years into the future, have tended to overproject the number of AIDS cases, sometimes by sizable amounts. There are two natural explanations for this pattern of forecast errors: (1) changes in the availability and use of drug therapies for the treatment of asymptomatic HIV infection (Gail, Rosenberg, and Goedert, 1990) and (2) changes in behaviors that are associated with the rate of occurrence of new infections. [13] Both explanations are examined empirically in the sections below, in the interest of improving the basis on which one can project AIDS cases.

##### A. Effects of AZT Usage

It is now generally believed that the use of AZT delays the onset of AIDS symptoms among HIV-infected individuals (i.e., "the risk of progression to AIDS among treated patients is approximately one-third the risk for untreated patients"; Centers for Disease Control, 1990, p. 15). The use of AZT as a prophylactic against the progression to AIDS began around 1987. In that year, approximately 7 percent of HIV-seropositive non-AIDS members of the Multicenter AIDS Cohort Study received some drug therapy to inhibit the onset of AIDS symptoms. In 1989, 73 percent of HBM in the San Francisco Clinic Study who would most benefit from AZT treatment received AZT (see Centers for Disease Control, 1990 and Hellinger, 1990). Although the increase from 7 to 73 percent from 1987 to 1989 likely overstates the true rise in the rate of increase in the use of AZT for the seropositive population at large (because HBM have relatively good access to knowledge about the latest prophylactic therapies and because the latter estimate focuses on a sample of highly-selected individuals), it is useful to explore the implications of these statistics for the time-series pattern of AIDS cases. [14]

Our analysis involves (1) converting the Weibull incubation distribution reported in Gail and Brookmeyer (1988) into a hazard function, introducing the effect of AZT by (2) constructing a new hazard function for the progression to AIDS that is one-third the level of that reported in Gail and Brookmeyer (1988); (3) working backwards from that new hazard function to the implied density function for the delay between infection and symptoms (see Figure 5 for a comparison of the pre- and post-AZT density functions for the delay between infection and

symptoms); (4) applying the modified density function to the time-series of newly-infected individuals estimated by Taylor (1989), for the fraction of individuals assumed to be using AZT (linearly interpolated for 1988); and (5) comparing the time-series of future AIDS cases (forward-calculated) between the two incubation distributions.

The results of this analysis are displayed in Figure 6. They indicate that the use of AZT by seropositive asymptomatic individuals could have accounted for a substantial short run reduction in the growth rate of the number of AIDS cases. This reduction should be interpreted as a likely upper bound on the true decrease since it is based on estimates of the level and growth of AZT usage that are likely upward biased.

Our results also indicate that the effect of AZT is to postpone cases, giving rise to a notch in the time-series of cases (i.e., the leveling that starts in 1987 followed by a steepening about two years later). Although the use of AZT might help explain the overprojections following 1986, the timing pattern suggested in Figure 6 is not readily apparent in the time series plots in Figures 1 and 2 of either all cases or definitionally consistent cases. This last result is not sufficient to reject the importance of AZT to the generation of AIDS case data because of the possible presence of other confounding influences (e.g., changes in new infection rates, changes in the pattern of AZT usage over time, particular instabilities in the distribution of reporting delays, definitional changes, or AZT not having such a large effect or a different pattern of effect at each stage, etc.).

To explore further the importance of AZT, we can reexamine the data for a particular transmission category: HBM. The results, which are displayed in Figure 3a, are quite striking insofar as the leveling of new cases appears to have predated the widescale use of AZT by as much as two years. This finding does not support the importance of AZT usage as an explanation for overprojections of the incidence of new AIDS cases. [15]

While there is little evidence that the use of AZT has affected the time series of AIDS cases, there is some evidence that AZT and other prophylaxis, particularly aerosolized pentamidine (which was introduced at about the same time as AZT), have had an impact on the survival time of AIDS patients. This finding is apparent in a comparison of death rates over time between the HBM and IVDU transmission categories (see Figures 7a and 7b). The series for HBM shows a marked flattening between 1987 and early 1988, when AZT came into heavy use, unlike the steady growth of deaths among IVDUs, who have a considerably lower usage rate of AIDS drugs than HBM (see Centers for Disease Control, 1990).

## B. Effects of Changes in High-Risk Behavior

AIDS projections based on empirical extrapolation all rely on the time-series of diagnosed AIDS cases reported to the CDC, adjusted for reporting delay. This time series embodies information on behavioral change, but with a long and variable lag because the mean of the

incubation distribution is over eight years. However, independent information on changes in behavior is available and can possibly be used to improve the accuracy of AIDS projections.

To illustrate the use to which such information can potentially be put, we propose a method that uses independent data on the time series of HIV infections. The empirical strategy is to apply an AIDS incubation distribution to these infection data to develop an index that reflects the time pattern of AIDS cases. Regression parameters are then estimated that link this index to the time series of actual AIDS cases. These parameters can then be applied to the values of the AIDS case index in future periods to project future AIDS cases. We implement this method using readily available and well-known data for two risk groups.

The infection data used in this illustration are for cohorts of individuals whose serostatus has been monitored since before the beginning of the epidemic and who have been tested annually for HIV infection. The two longest-running studies are the San Francisco Cohort Study of Gay Men and the Hershey Hemophilia Cohort Study. [16] The time-series of new infection rates for each cohort are reported in Table 4. These data indicate that the rate of new infection declined sharply in and after 1983 for HBM; the results for hemophiliacs show that HIV infection had saturated the population by 1985, after which the U.S. blood supply was effectively screened for the HIV. These patterns in rates of new infection would not manifest themselves significantly in patterns of new AIDS cases until the second half of the 1980s. Failure to account for this decline could lead to sizable projection errors.

The information on rates of new infection in Table 4 is used to construct an AIDS Case Index (ACI) by using the incubation distribution (i.e., the Weibull distribution in Figure 5, without AZT) to forward calculate AIDS cases from HIV infections. These indexes are reported in columns 2 and 4 in Table 4.

Before examining the results of analyses based on these indexes, their limitations should be noted. First, the degree to which the San Francisco cohort is representative of the U.S. population of HBM is uncertain. The San Francisco cohort is likely to have been better informed about the risks of different behaviors, available treatments, etc. In addition, the epidemic began earlier in San Francisco than in the rest of the U.S. Notwithstanding these differences, the post-1983 rates of new infection among HBM cohorts in San Francisco are generally near the median of corresponding rates for other HBM cohorts elsewhere in the U.S. (Centers for Disease Control, 1987).

Second, the degree to which the Hershey Hemophilia cohort is representative of the entire U.S. population of users of blood and blood products is also questionable, mainly because the Hershey cohort is exceedingly small in size and is limited to severe hemophiliacs. Also, the U.S. blood supply began to be informally screened (with blood banks asking prospective blood donors about behavior that might have put them at high risk for contracting the HIV) as early as 1983, which could have affected the rate of new infection due to blood transfusions. However, this type of informal screening would not have reduced to the same extent the probability of infection

among severe hemophiliacs who received clotting factor derived from the blood of many individuals.

Third, using the ACI to project AIDS cases more than two years into the future requires one to make an assumption about the future time series of new infections. In the example below, we have assumed a stable rate of new infection for HBM of 0.8 percent per year.

These potential limitations of the AIDS Case Indexes in Table 4 can be assessed empirically by seeing how closely the various historical time series' of AIDS cases have moved with these indexes. Table 5 reports time-series regressions of the number of AIDS cases (adjusted by the CDC for reporting delay) on the appropriate ACI. Estimates are reported separately for (1) the time periods 1982-86 and 1982-90, (2) HBM and BTBP cases, and (3) diagnosed cases and definitionally consistent cases. [17]

The results in Table 5 are consistent with the hypothesis that the ACI contains useful information about the time series pattern of AIDS cases. All of the index coefficients are positive and the R-squared estimates are quite high (and higher in magnitude than those that result from corresponding time-series regressions of AIDS cases on a linear time trend). Also, in contrast to the quadratic trend regressions in Table 3, in which the coefficients changed sizably and significantly between sub-periods, the coefficients appear remarkably stable between 1982-86 and 1982-90, an indication that the ACI models are appropriate. When we allow a linear time trend and the appropriate ACI to "fight it out" in the same regression specification over the 1982-90 period, the trend coefficient always falls sharply in magnitude (sometimes changing sign) while the ACI coefficient exhibits relatively modest changes (always remaining positive). These findings suggest that the ACI model is reasonably robust with respect to a simple change in specification. [18]

Figures 3a, 3b, 4a, and 4b plot (1) CDC data on the time series of HBM and BTBP AIDS cases (available monthly) and HBM and BTBP consistent cases (available quarterly) from 1982 to 1990; (2) projected cases in each category based on a time-series regression of case counts on a quadratic trend; and (3) fitted and projected cases based on a time-series regression of case counts on the ACI from 1982 to 1986 and projected using the ACI from 1987 to 1990.

The results in Figures 3a, 3b, 4a, and 4b provide a strong indication that AIDS projections can be dramatically improved by exploiting knowledge about the process generating the data: that HIV infections (about which some information is available) lead to AIDS cases according to a lag distribution (about which some information is also available). For HBM cases, the projection errors for 1990 that result from a standard extrapolation model are four to eight times larger (in absolute value) than those that emerge from the ACI model; for BTBP cases they are five to eight times larger. [19]

## V. Conclusion

To illuminate the potential importance of our empirical findings, we have constructed several crude projections of numbers of new AIDS cases in 1994 and 1995. These projections, which refer to diagnosed cases that will ever be reported to the CDC, are presented in Table 6. The projections in the first panel are based on a quadratic extrapolation model applied to monthly CDC data on all AIDS cases from 1982 through 1990 (adjusted for reporting delay by the CDC and for the September 1987 case definition change by the inclusion of a shift parameter, but not adjusted for underreporting). The projections in rows two through four are derived by fitting a similar model to case data for the three transmission categories indicated. The fifth row is a projection of cases due to other forms of transmission, derived by subtracting projected HBM, BTBP, and IVDU cases from the projection for All Cases. [20]

The projections in the second panel are based on the AIDS Case Index in Table 4, constructed separately for HBM and BTBP cases. The final row in Table 6 reports a Composite (ACI-Extrapolation) Projection, constructed by adding the ACI projections for HBM and BTBP Cases to the extrapolation-based projections for IVDU Cases and Other Cases (since we have not constructed an ACI for either of these latter transmission categories and since Figure 3c suggests that the extrapolation method results in relatively small projection errors when fit to IVDU cases).

There are two noteworthy features of the projections in Table 6. First, the standard extrapolation method projects a 72 percent increase in the number of new cases between 1991 and 1995 (with a 10 percent increase from 1994 to 1995). By contrast, the composite projection suggests only a 24 percent increase in AIDS cases from 1991 to 1995 (with a four percent increase from 1994 to 1995). The extrapolation method projects 22,000 more new cases in 1995 than the composite method.

Second, the extrapolation method projects HBM as the dominant AIDS transmission category in 1994 and 1995, whereas IVDU will be the dominant category according to the composite projection. This difference has important policy ramifications with respect to the control of the epidemic insofar as it suggests that efforts aimed at the prevention, detection, and treatment of HIV/AIDS need to be focussed on a fundamentally different social and economic population, concentrated in different areas of the U.S. This result also highlights the finding that behavioral change among homosexual/bisexual males in the U.S. seems to have been substantial enough to dramatically alter the course of the epidemic within that community.

Our composite projections could be further refined by utilizing other sources of information on (or correlated with) rates of new infection among HBM and BTBP, adjusting available infection data to improve the degree to which they represent the entire HBM and BTBP populations, and gathering data related to rates of new infection among IVDU and other transmission categories (even for relatively small samples of individuals), and for specific demographic groups and geographic locations (see Bloom and Glied, 1992). Nonetheless, our analysis has shown that even the crude ACI method we use to project new AIDS cases

outperforms the extrapolation method, as judged by historical data for two transmission categories. The figures in Table 6 show further that different projection methods can lead to substantially different projections, even as few as four to five years into the future. Taken together, these findings confirm the importance of paying attention to the process generating the data series for which projections are desired.

The literature on AIDS projections has exhibited a preoccupation with statistical issues related to adjustments for underreporting, reporting delay, and changes in the AIDS case definition, and to the choice of functional form in extrapolation equations and the imposition of non-negativity constraints in back calculation. Although we do not deny the importance of these and other statistical issues, our results suggest that relatively greater improvements in projection accuracy can be realized by attention to underlying behavioral processes and through the incorporation of additional readily available data into AIDS case projections. Our results suggest that further application of the ACI model can dramatically improve the accuracy of the next generation of AIDS projections and the basis on which individual behavior and related public and private policies are determined.

## Notes

1. But see Lemp, Payne, and Rutherford, 1990 for an example of a micro- simulation model that provides highly accurate AIDS case projections for San Francisco.
2. See Harris, 1990 for an in-depth analysis of reporting delays and changes in their distribution over time.
3. See Centers for Disease Control (1989) for a review of changes that occurred in the AIDS case definition prior to 1987. Note also that another major definitional change due to have taken effect in January 1992 was postponed indefinitely to allow for further study of the impacts of the proposed change.
4. Diagnosed cases from 1991 were estimated by examining the ratio of cases diagnosed and reported in 1987 to cases diagnosed in 1987 and reported by the end of 1991. That ratio, 55 percent, was then applied to cases diagnosed and reported in 1991 to estimate the number of 1991 diagnosed cases that would ultimately be reported by the end of 1995. Other calculations we performed suggest this is a reasonable estimation procedure.
5. The low end of this projection range, the Brookmeyer and Damiano lower bound projection, is based on three critical assumptions: that no new infections occur after 1987, that AZT becomes very widely used by 1991 among asymptomatic HIV-infected individuals, and that the use of AZT delays the onset of symptoms among seropositive individuals who are asymptomatic. However, there is considerable evidence that HIV infections continued to occur after 1987 and that AZT is not widely used by all groups of asymptomatic HIV-infected individuals (Centers for Disease Control, 1990). Thus, the Brookmeyer and Damiano lower bound projection comes closest among the studies in Table 2 to predicting the actual number of new cases in 1991, but not because the assumptions upon which that projection is based are closely satisfied.
6. The Hay and Wolak (1990) study utilizes data from before and after the AIDS case definition changed in September 1987, without adjusting for the effect of the definitional change on the underlying data used to construct their projections. According to Gail and Brookmeyer (1988) this definitional change, which mainly involved a weakening of the criteria that had to be satisfied for a physician to render a diagnosis of AIDS, caused roughly an 18 percent increase in the number of AIDS cases that would ever be reported to the CDC. The definitional change presumably affected the timing of AIDS case reports as well, resulting in a temporary upsurge of cases in the period immediately following the change. Despite this limitation, Hay and Wolak's projection of 1991 cases (44,000) comes very close to our estimate of diagnosed cases for that year (46,000).
7. We are grateful to John M. Karon of the CDC for providing us with these data, which also embody the CDC adjustment for reporting delay. Note that consistent cases do not refer to either the AIDS case definition in effect before or the definition in effect after the September 1987

definitional change; rather, they refer to a distinctly different definition that can be consistently applied during both time periods.

8. Brookmeyer and Damiano report projections of cumulative AIDS cases through the end of 1991; Morgan and Curran report projections of new cases in 1991.

9. The HBM category excludes cases in which intravenous drug use is also a risk factor; the IVDU category does not include any cases in which HBM is a risk factor.

10. The fact that the leveling off began later in the monthly case data than in the quarterly data on consistent cases supports the notion that the case definition change had an effect on the timing of cases.

11. This result is also upheld using 1986 or 1988 as the break points.

12. We also explored the use of the following logistic growth curve to model the time series of cases:  $Y_t = a/(1 + \exp(c + d^t))$ . Although this curve provides a reasonably good fit to data on all cases, as well as cases by transmission category, it does not perform as well as the quadratic specification in out-of-sample projections. The results are available from the authors upon request.

13. This conclusion assumes reasonably stable patterns of underreporting and reporting delays and that the effects of case definition changes have been accounted for adequately.

14. For other discussions of the effect of using AZT and its impact on the pattern and count of AIDS cases see Brookmeyer and Liao, 1990; Hellinger, 1990; Gail and Brookmeyer, 1990; and Gail, Rosenberg, and Goedert, 1991.

15. Even if AZT has a small aggregate effect, failure to account for its use could lead to large errors in projections based on back calculation because of the sensitivity of that method to estimates of the incubation distribution.

16. Both cohort studies are discussed in Centers for Disease Control, 1987.

17. Since they map the ACI, which applies to a sample of known size, into numbers of AIDS cases in a population of unknown size (that is assumed to be fixed), the regression coefficients in Table 5 could, in principle, be used to estimate the sizes of the relevant transmission populations. The validity of this technique rests on the assumption that the prevalence of HIV infection at each point in time is equal between the sample and the population. This assumption of equality is not, however, necessary for the ACI to generate accurate projections of AIDS case counts. Rather, it is necessary only that the ratio of new infections in the sample and the population is roughly constant over time. Our results suggest that the prevalence of infection in the San Francisco cohort and in the Hershey Hemophilia cohort are greater than the prevalence of infection in the corresponding HBM and BTBP populations of the U.S.

18. Unfortunately, the small number of data points, due to the fact that the index is only available on an annual basis, precludes our constructing a meaningful test for the stability of the coefficients. The small number of data points also precludes our performing the necessary corrections for serial correlation that would enable us to report significance tests for the trend and ACI coefficients.

19. The ACI method of projection is far less sensitive to estimates of the incubation distribution than the method of back calculation because it uses that distribution only once in constructing projections. For example, applying the ACI method to the hypothetical example in Table 1 yields a projection of 5500 cases for time period 9 using the original incubation distribution and a projection of 5200 cases using the perturbed distribution (an (average) deviation of less than 6 percent, in contrast to the nearly 40 percent deviation for projections based on back calculation).

20. We have not constructed projections based on the method of back calculation because 1994 and 1995 are too far into the future to generate meaningful projections using this method.

## References

- Becker, Marshall, and Jill Joseph. 1988. "AIDS and Behavioral Change to Reduce Risk: A Review." American Journal of Public Health 78(4): 394-410.
- Bloom, David E. and Sherry Glied. 1992. "Projections of the Number of New AIDS Cases in the U.S." Mimeo.
- Brookmeyer, Ron. 1991. "Reconstruction and Trends of the AIDS Epidemic in the United States." Science 253: 37-42.
- Brookmeyer, Ron, and Anne Damiano. 1989. "Statistical Methods for Short Term Projections of AIDS Incidence." Statistics in Medicine 8 (January): 23-34.
- Brookmeyer, Ron and Mitchell Gail. 1986. "Minimum size of the Acquired Immunodeficiency Syndrome (AIDS) Epidemic in the United States." Lancet 83: 301-308.
- Brookmeyer, Ron, and J. Liao. 1990. "Statistical Modelling of the AIDS Epidemic for Forecasting Health Care Needs." Biometrics 46(4): 1151-63.
- Centers for Disease Control. 1991. Memo, Reprinted in AIDS Resource Guide. Atlantic Information Service, 1991 September.
- Centers for Disease Control. 1990. "HIV Prevalence Estimates and AIDS Case Projections for the United States: Report Based upon a Workshop." Morbidity and Mortality Weekly Report Nov. 30, 1990, vol. 39, RR-16.
- Centers for Disease Control. HIV/AIDS Surveillance Report. Various issues.
- Centers for Disease Control. 1989. "AIDS and Human Immunodeficiency Virus Infection in the United States: 1988 Update." Morbidity and Mortality Weekly Report May 12, 1989, vol. 38, S-4.
- Centers for Disease Control. 1987. "Human Immunodeficiency Virus Infection in the United States: A Review of Current Knowledge." Morbidity and Mortality Weekly Report Dec. 18, 1987, Vol. 36, S-6.
- Conway, G.A., B. Colley-Niemeyer, C. Pursley, et al. 1989. "Underreporting of AIDS Cases in South Carolina, 1986 and 1987." JAMA 262(20): 2859-63.
- Eden, Jill. 1988. AIDS and Health Insurance: an OTA Survey. Office of Technology Assessment, Washington, Feb. 1988.

Gail, Mitchell, and Ron Brookmeyer. 1988. "Methods for Projecting the Course of the Acquired Immunodeficiency Syndrome Epidemic." Journal of the National Cancer Institute 80(12): 900-912.

Gail, Mitchell, P.S. Rosenberg, and J.J. Goedert. 1990. "Therapy May Explain Recent Deficits in AIDS Incidence." Journal of AIDS 3(4): 296-306.

Hardy, Ann M., T.E. Starcher, and W. Meade Morgan. 1987. "Review of Death Certificates to Assess Completeness of AIDS Case Reporting." Public Health Reports 102: 388-391.

Harris, Jeffrey E. 1990. "Reporting Delays and the Incidence of AIDS." Journal of the American Statistical Association, 85(412): 915-924.

Hay, Joel and F. Wolak. 1990. "Bootstrapping HIV/AIDS Projection Models: Back Calculation with Linear Inequality-Constrained Regression." Hoover Institution Working Paper in Economics, No. E-90-5.

Hellinger, Fred J. 1990. "Forecasting the Number of AIDS Cases: An Analysis of Two Techniques." Inquiry 27(Fall): 212-224.

Hellinger, Fred J. 1988. "Forecasting the Personal Medical Care Costs of AIDS from 1988 through 1991." Public Health Reports 103: 309-319.

Hendrix, Omar L. 1987. "New York City Health and Hospitals Corporation", Ch. 14 in AIDS: Public Policy Dimensions (John Griggs, editor). United Hospital Fund, New York.

Karon, John M., O.J. Devine, W.M. Morgan. 1989. "Predicting AIDS Incidence by Extrapolating from Recent Trends", pp. 58-88 in Mathematical and Statistical Approaches to AIDS Epidemiology (C. Castillo-Chavez, editor). Lecture Notes in Biomathematics, vol. 83, Berlin: Springer Verlag.

Karon, John M., Timothy Dondero, Jr., James W. Curran. 1988. "The Projected Incidence of AIDS and Estimated Prevalence of HIV Infection in the United States." Journal of Acquired Immune Deficiency Syndrome 1 (6): 542-550.

Lemp, George, Susan Payne, and George Rutherford. 1990. "Projections of AIDS Morbidity and Mortality in San Francisco." JAMA 263(11): 1497-1501.

MetLife. 1990. "AIDS: MetLife's Experience 1986-1989", MetLife Statistical Bulletin, Oct-Dec 1990: 2-9.

Morgan, W. Meade, and James W. Curran. 1986. "Acquired Immunodeficiency Syndrome: Current and Future Trends." Public Health Reports, 101(5): 459-465.

Public Health Service. 1986. Coolfont report: A PHS Plan for Prevention and Control of AIDS and the AIDS Virus. Public Health Reports, 101: 341-8.

Rayhawk, Peggie. 1991. "National Survey of State-Only HIV/AIDS Funding". Intergovernmental AIDS Reports, November.

Rhein, Reginald. 1990. "AIDS Research Funds to Hit a Plateau". Journal of NIH Research, June: 21.

Taylor, J.M. 1989. "Models for the HIV Infection and AIDS Epidemic in the U.S." Statistics in Medicine 8(1): 45-48.

United States General Accounting Office. 1989. "AIDS Forecasting: Undercount of Cases and Lack of Key Data Weaken Existing Estimates". GAO Report, June 1989.

Volberding, Paul A.. 1987. "San Francisco General Hospital", Ch. 15 in AIDS: Public Policy Dimensions (John Griggs, editor). United Hospital Fund, New York.

Table 1

Illustrative Example of Back Calculation Method Used to Project the Incidence of AIDS One Period into the Future (Based on Purely Hypothetical Information)

time period	AIDS Cases observed/ (projected)	Expressed as a Distributed Lag on the Number of Prior Infections (I)	Inferred HIV Infections	(Assumed)
9	(5500)	$.1(I_6) + .4I_5 + .5I_4$	n.i.	
8	2700	$.1I_5 + .4I_4 + .5I_3$	n.i.	
7	950	$.1I_4 + .4I_3 + .5I_2$	n.i.	
6	300	$.1I_3 + .4I_2 + .5I_1$	n.i.	(0)
5	50	$.1I_2 + .4I_1 + .5I_0$	10,000	
4	0	$.1I_1 + .4I_0 + .5I_{-1}$	3,000	
3	...	...	1,000	
2	...	...	500	
1	...	...	0	

1. This illustrative example was constructed under the following assumptions about the delay between HIV infection and the development of AIDS:

50 percent of individuals infected in period  $t$  develop AIDS in period  $t+5$ ;

40 percent of individuals infected in period  $t$  develop AIDS in period  $t+4$ ;

10 percent of individuals infected in period  $t$  develop AIDS in period  $t+3$ .

2. The number of infections cannot be negative. This restriction implies that  $I_{-1}$ ,  $I_0$ , and  $I_1$ , are all equal to zero.

3. n.i. denotes cells for which the number of infections is not inferable from the observed true series of AIDS cases using the backcalculation algorithm.

4. Projections can also be constructed under alternative assumptions about the number of infections in cells marked n.i.

Table 2  
Projections of the Incidence of AIDS

Study (method)	Time Period of Data Used to Construct Projection	1991 Projection (in thousands)	1991 Projection Adjusted for 1987 Definition Change *	Method of Correcting for Reporting Delay	Definition of an AIDS Case used for Projection
Brookmeyer, Liao, 1980 (1) (back calculation)	quarterly 1/80-7/87 (reported by 10/89)	45 - 75(2)	53-89	non-parametric estimate of delay distribution	pre-1987 definition
Hay, Wolak, 1990 (3) (back calculation)	1/79-12/89	44	not applicable	estimated delay distribution	not adjusted for 1987 definition change
Morgan, Curran, 1986 (4) (extrapolation)	all cases reported 6/81-5/86	74	87	adjusted for delay, assuming stable pattern	pre-87 definition
Karon, Dondero, Curran, 1988 (5) (extrapolation)	7/83-6/87 (reported by 3/88)	64	69 (6)	adjusted for delay using CDC weights	consistent case definition
Hellingger, 1988 (7) (extrapolation)	1/84-10/87	51-74(8)	60-87	ratio of diagnosed to reported cases	pre-87 definition

Notes:

- (1) Gill and Brookmeyer (1988) estimate that 18 percent of the cases reported under the post-1987 definition would not have been reported under the CDC's pre-1987 definition. For those studies that used pre-1987 data, the 1991 estimates have been inflated by 18 percent to account for the definition change.
- (2) Brookmeyer, Ron, and J. Liao, 1990. "Statistical Modelling of the AIDS Epidemic for Forecasting Health Care Needs." *Biometrics* 46(4): 1151-63.
- (3) Both estimates assume no new infections after 1987. Lower estimate assumes that 75 percent of infected individuals with CD4 counts of less than 200 were using AZT in 1991. Hay, Joel and F. Wolak, 1990. "Bootstrapping HIV/AIDS Projection Models: Back Calculation with Linear Inequality-Constrained Regression." Hoover Institution Working Paper in Economics, No. E-90-5.
- (4) Morgan, W. Meade, and James W. Curran, 1986. "Acquired Immunodeficiency Syndrome: Current and Future Trends." *Public Health Reports*, 101(5): 459-465.
- (5) Karon, J.M., Timothy Dondero, Jr., James W. Curran, 1988. "The Projected Incidence of AIDS and Estimated Prevalence of HIV Infection in the United States." *Journal of Acquired Immune Deficiency Syndrome* 1 (6): 542-550.
- (6) Karon, Dondero, and Curran estimate that the consistent case measure is roughly nine percent lower than the post-1987 definition measure. The adjusted projection column for this study, therefore, been adjusted correspondingly.
- (7) Hellingger, Fred J. 1988. "Forecasting the Personal Medical Care Costs of AIDS from 1988 through 1991." *Public Health Reports*, 103:309-319.
- (8) Range reflects alternative extrapolation specifications.

**Table 3**  
**Results of Extrapolation Projections**  
 (all figures adjusted for reporting delay using CDC weights)

	<u>Intercept</u>	<u>Time</u>	<u>Time<sup>2</sup></u>	<u>R<sup>2</sup></u>	<u>N</u>	<u>F-statistic</u> *
Quadratic Specification:						
Diagnosed Cases (monthly)						
All Cases 82-86	87.91	-3.12	0.52	.99	60	
All Cases 82-90	-305.62	27.84	0.13	.97	108	104.0
HBM Cases 82-86**	47.58	-1.26	0.32	.99	60	
HBM Cases 82-90	-238.66	22.12	0.00	.96	108	116.8
BTBP Cases 82-86***	2.78	-0.38	0.02	.95	60	
BTBP Cases 82-90	-25.53	1.02	-0.01	.87	108	94.7
IVDU Cases 82-86	17.79	-1.02	0.10	.98	60	
IVDU Cases 82-90	-44.10	3.09	0.06	.97	108	72.6
Consistent Cases (Quarterly)****						
All 83-86	435.29	61.96	13.89	.99	16	
All 83-90	-562.98	356.97	-0.77	.98	31	46.1
HBM 83-86	287.46	42.22	9.07	.99	16	
HBM 83-90	-426.4	261.39	-1.88	.98	31	46.7
BTBP 83-86	7.24	1.37	0.64	.97	16	
BTBP 83-90	-54.03	20.05	-0.32	.93	31	20.2
IVDU 83-86	78.12	12.11	2.46	.99	16	
IVDU 83-90	-60.79	47.62	1.02	.99	31	30.3

\* The F-statistic is reported as a test of the hypothesis that the coefficient vector is stable before and after 1987. The critical value at the 95 percent confidence level for monthly cases is 3.98; the critical value at the 95 percent confidence level for quarterly cases is 4.68. The unconstrained results are available from the authors upon request.

\*\* Homosexual/bisexual men who are not intravenous drug users.

\*\*\* Blood transfusion/blood products cases.

\*\*\*\* Consistent case estimates are for CDC data adjusted to include only cases that would have met the pre-1987 case definition. These data do not include children, Pattern II, and "No Identified Risk" cases. (Karon, 1991). These data end in the third quarter of 1990.

Quadratic specification  $Y_t = a + bT + cT^2$

Table 4  
Construction of Indices(1)

	HBM Transmission*		BTBP Transmission*	
	Infection (2) <u>Distribution</u>	AIDS (3) <u>Case Index</u>	Infection <u>Distribution</u>	AIDS <u>Case Index</u>
1978	1.1	0	0	0
1979	2.8	0	10	0
1980	11	0.02	10	0
1981	14.8	0.1	6.7	0.2
1982	20.8	0.39	40	0.6
1983	2.1	0.95	23.3	1.0
1984	2.1	1.85	10	2.3
1985	2.1	2.75	0	4.0
1986	0.8	3.58	0	5.5
1987	0.8	4.51	0	7.0
1988	0.8	5.18	0	8.4
1989	0.8	5.49	0	9.4
1990	0.8	5.49	0	9.6
1991	0.8	5.31	0	9.4
1992	0.8	4.88	0	8.8
1993	0.8	4.41	0	7.8
1994	0.8	3.83	0	6.7
1995	0.8	3.2	0	5.6

\* Homosexual/bisexual men who are not intravenous drug users.

\*\* Blood transfusion/blood products cases.

(1) The incubation distribution used to construct the AIDS case index is based on Brookmeyer and Gail's (1989) estimate of the cumulative distribution for the transition to AIDS:  $F(t) = 1 - \exp(-0.004t^{2.438})$  where  $t$  represents the delay (in years) between infection and the emergence of AIDS symptoms and  $F(t)$  is the cumulative distribution function for a Weibull variate.

(2) The infection distribution represents the percent of initially seronegative individuals in a well-defined cohort who became infected in each calendar year. The homosexual/bisexual infection distribution uses information from a San Francisco cohort consisting of 283 individuals. The blood products infection distribution uses information from the Hershey Hemophilia cohort of 30 individuals. (CDC, 1987) Since the infection distribution reported in CDC, 1987 for homosexual/bisexual men ends in 1986, we have assumed that it remained constant at 0.8 persons per year thereafter. Heat treatment of blood products and testing of the blood supply were implemented in 1985, implying few BTBP infections after 1985. In this connection, note that the Hershey cohort was 100 percent infected by the end of 1984.

(3) The case indices are constructed by applying the incubation distribution to the infection rate in each year and then summing AIDS cases across infection cohorts to estimate a case index by year.

**Table 5**  
**Results from Index Based Projections**  
 (see Table 4 for construction of index)

	1982-86	HBM <sup>*</sup> 1982-90	1982-86	BTBP <sup>**</sup> 1982-90
	Diagnosed Cases (diagnosed in July of each year)			
Intercept	-130.93	-236.87	-7.22	-4.63
AIDS Case Index	301.82	370.80	10.47	10.58
R <sup>2</sup>	.96	.98	.92	.85
N	5	9	5	9
	Consistent Cases <sup>***</sup> (diagnosed in third quarter of each year)			
Intercept	-584.91	-830.58	-29.49	-9.6
AIDS Case Index	977.07	1097.20	33.67	28.47
R <sup>2</sup>	.98	.99	.96	.92
N	4	8	4	8

\* Homosexual/bisexual men who are not intravenous drug users.

\*\* Blood transfusion/blood products cases.

\*\*\* Consistent case estimates are for CDC data adjusted to include only cases that would have met the pre-1987 case definition. These data do not include children, Pattern II, and "No Identified Risk" cases. (Karon, 1991). These data end in the third quarter of 1990.

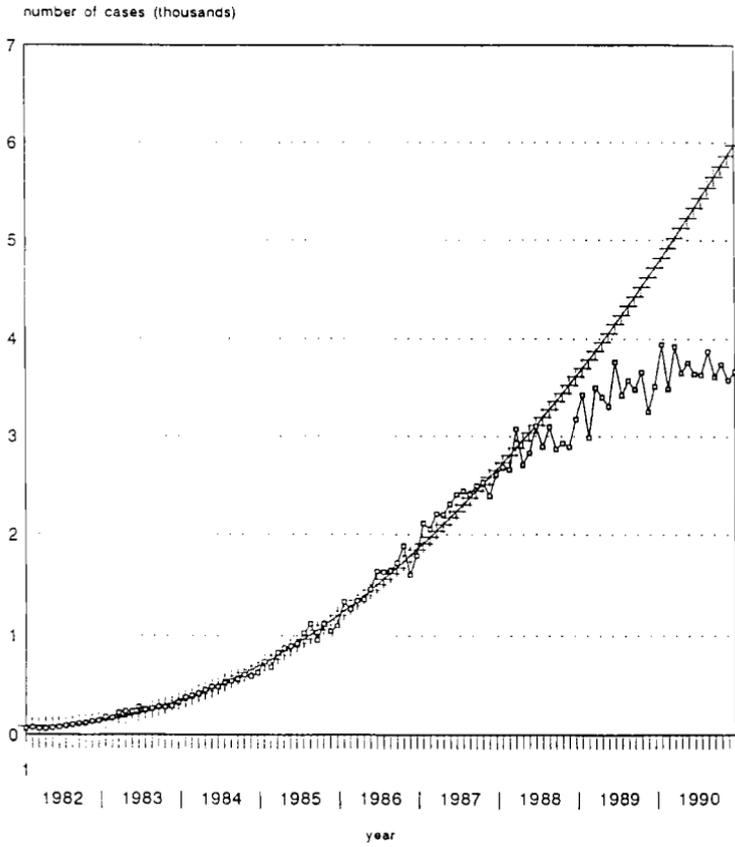
TABLE 6

Selected Projections of AIDS Cases  
in 1994 and 1995

<u>Extrapolation Method<sup>1</sup></u>	<u>1994</u>	<u>1995</u>
All Cases	72,000	79,000
HBM Cases	32,000	34,000
BTBP Cases	1,000	-0-
IVDU Cases	20,000	23,000
Other Cases <sup>2</sup>	19,000	22,000
 <u>ACI Method<sup>3</sup></u>  		
HBM Cases	15,000	11,000
BTBP Cases	1,000	1,000
 <u>Composite Projection</u>  		
All Cases <sup>4</sup>	55,000	57,000

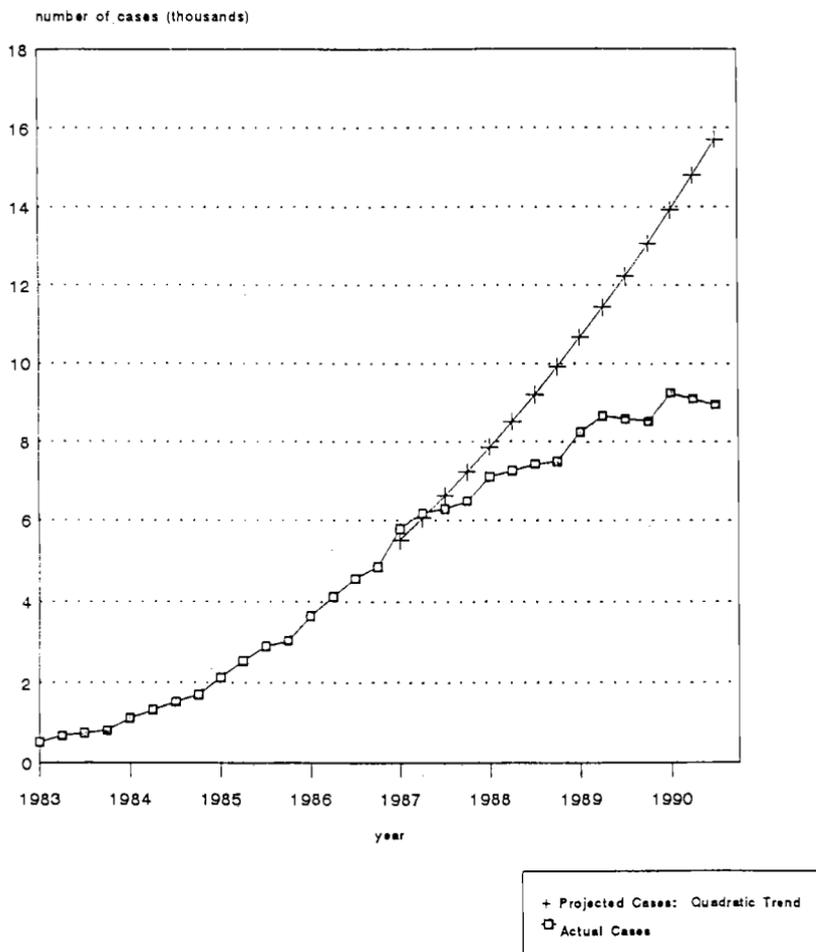
- 
- 1 The extrapolation estimates for All Cases, HBM Cases, BTBP Cases, and IVDU Cases are based on the estimated coefficients of a quadratic trend fit to monthly data for 1982-90 as reported in Table 3, with an intercept shift following the change in the AIDS case definition that occurred in September, 1987. All figures in this table are rounded to the nearest thousand.
  - 2 The estimate for Other Cases is equal to the difference between the extrapolation estimate for All Cases and the sum of the extrapolation estimates for HBM, BTBP, and IVDU Cases.
  - 3 The ACI estimates for HBM and BTBP Cases are calculated by applying the estimated coefficients from the ACI fit to monthly data for 1982-90 (reported in Table 5) to the ACI values for 1994 and 1995 for HBM and BTBP Cases (reported in Table 4).
  - 4 The Composite Projection is constructed by adding the ACI projection of HBM and BTBP Cases to the extrapolation projection for IVDU and Other Cases.

Figure 1  
Monthly AIDS Cases  
(adjusted for reporting delay)



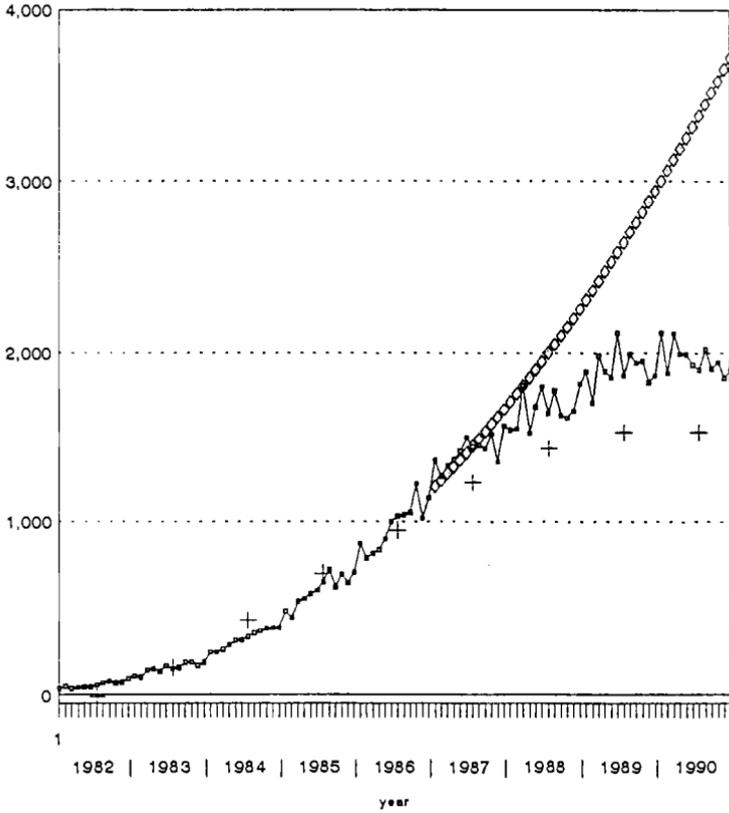
Source: Centers for Disease Control

Figure 2  
Quarterly AIDS Cases Based on Consistent Case Definition  
(adjusted for reporting delay)



Source: J.M. Karon, unpublished.

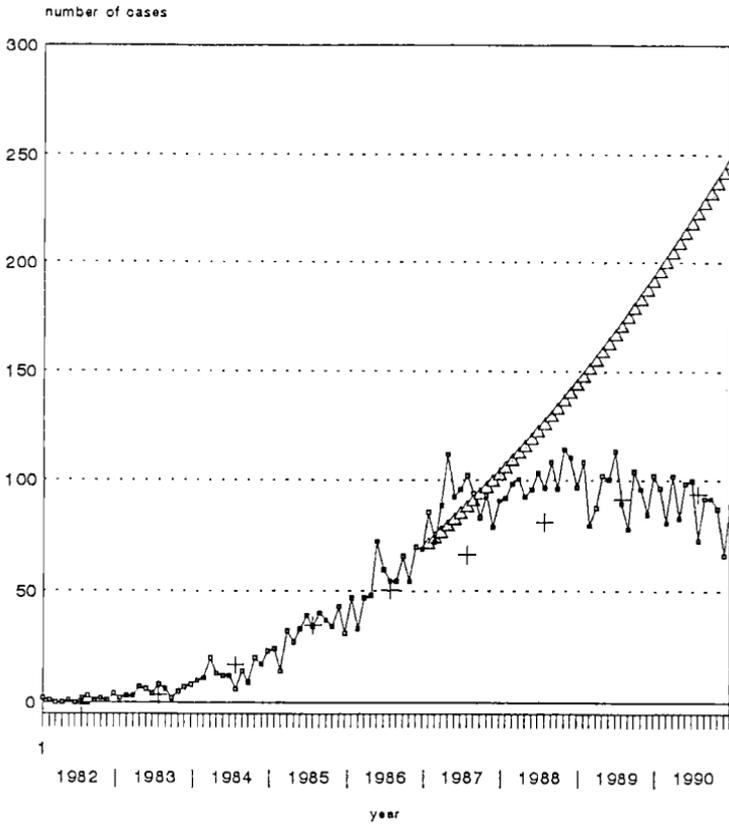
Figure 3a  
 Monthly AIDS Cases: HBM  
 (adjusted for reporting delay)



+ Projected HBM Cases: AIDS Case Index  
 ◆ Actual HBM Cases ◆ Projected HBM Cases\*

\* Quadratic trend  
 Source: Centers for Disease Control

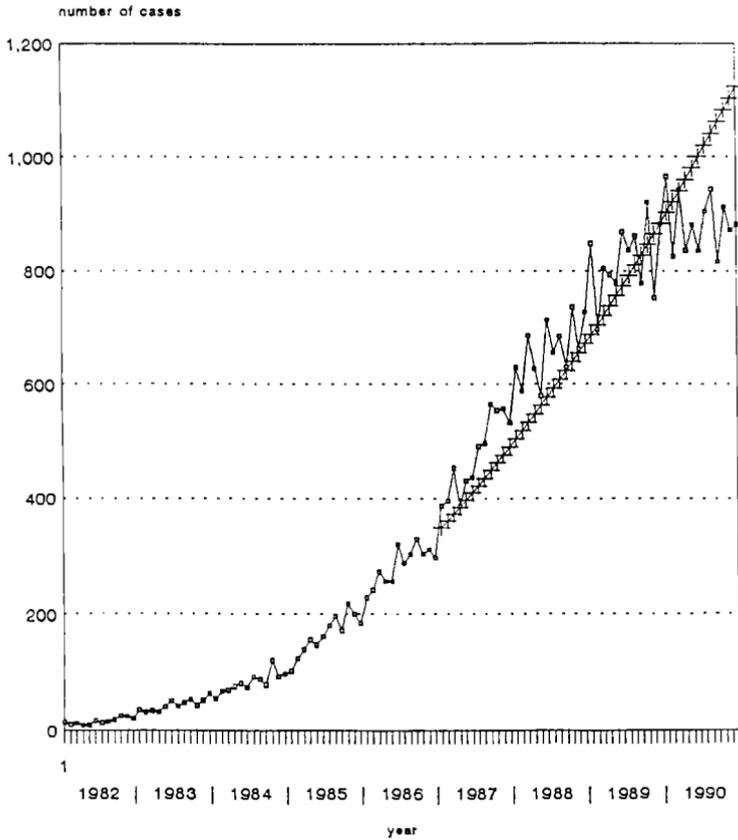
Figure 3b  
 Monthly AIDS Cases: BTBP  
 (adjusted for reporting delay)



+ Projected Cases: AIDS Case Index  
 ■ Actual BTBP Cases ▲ Projected BTBP Cases

\* Quadratic Trend  
 Source: Centers for Disease Control

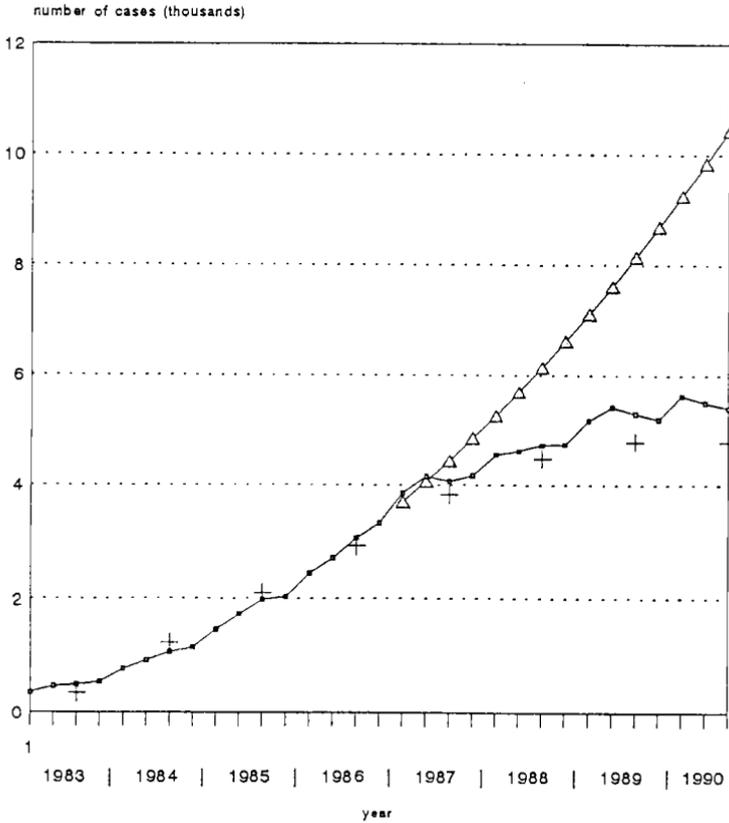
Figure 3c  
Monthly AIDS Cases: IVDU  
(adjusted for reporting delay)



+ Projected IVDU Cases: Quadratic Trend  
- Actual IVDU Cases

Source: Centers for Disease Control

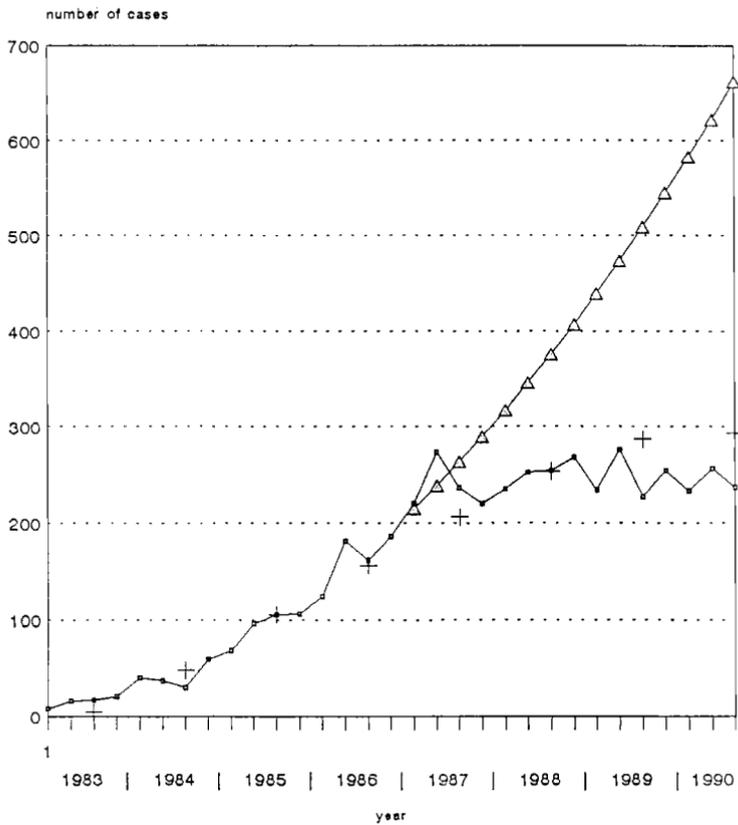
Figure 4a  
 Quarterly AIDS Cases Based on Consistent Case Definition: HBM  
 (adjusted for reporting delay)



+ Projected HBM Cases: AIDS Case Index  
 ■ Actual HBM Cases △ Projected HBM Cases\*

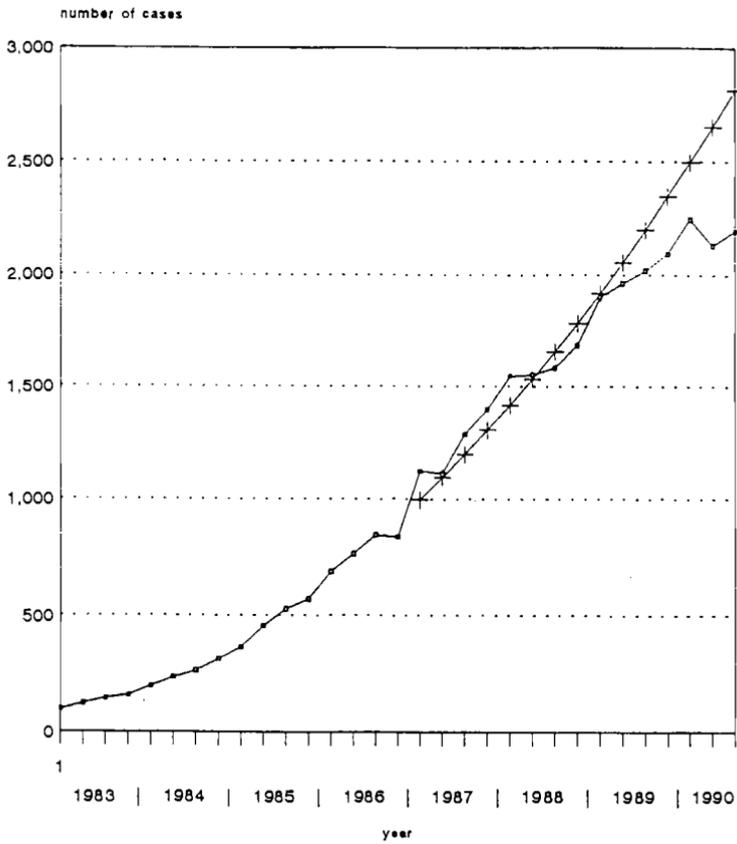
\* Quadratic Trend  
 Source: J.M. Keron, unpublished.

Figure 4b  
 Quarterly AIDS Cases Based on Consistent Case Definition: BTBP  
 (adjusted for reporting delay)



\* Quadratic Trend  
 Source: J.M. Karon, unpublished.

Figure 4c  
 Quarterly AIDS Cases Based on Consistent Case Definition: IVDU  
 (adjusted for reporting delay)



+ Projected IVDU Cases: Quadratic Trend  
 — Actual IVDU Cases

Source: J.M. Karon, unpublished.

Figure 5  
Distribution of Time Between HIV Infection and AIDS:  
Pre- and Post-AZT

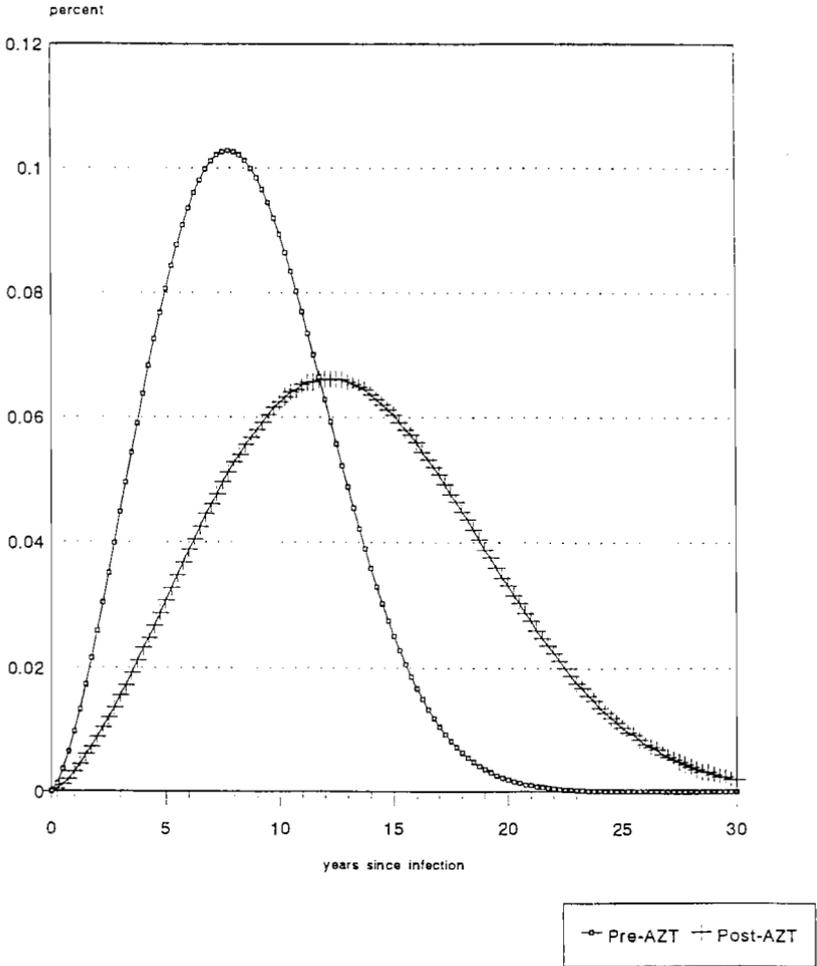
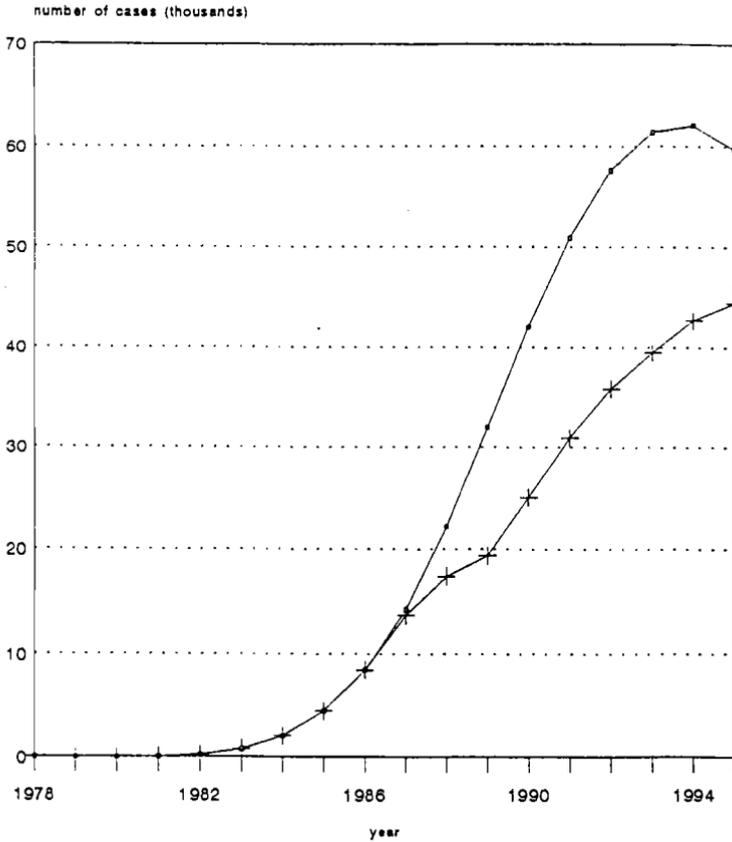
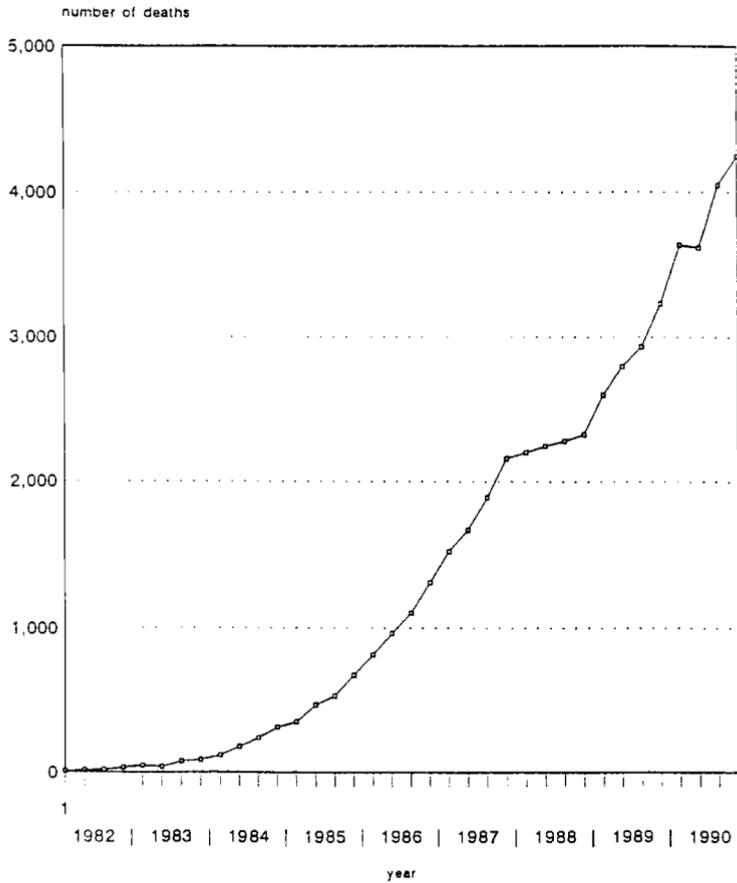


Figure 6  
Effect of AZT on the Time Series of AIDS Cases  
(assumes no new infections after 1987)



Expected time series:  
—•— pre-AZT —+— post-AZT

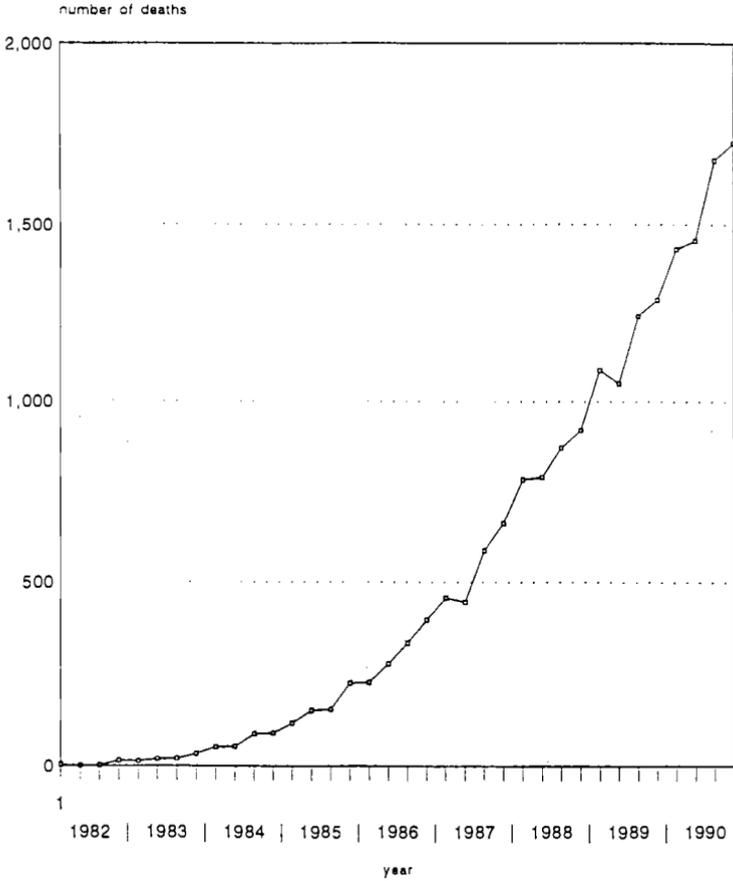
Figure 7a  
Quarterly HBM Deaths



◻ HBM Deaths

Source: Centers for Disease Control

Figure 7b  
Quarterly IVDU Deaths



—○— IVDU Deaths

Source: Centers for Disease Control