Impact of Epidemic and Individual Heterogeneity on the Population Distribution of Disease Progression Rates

An Example from Patient Populations in Trials of Human Immunodeficiency Virus Infection

John P. A. Ioannidis,1 Joseph C. Cappelleri,2 Christopher H. Schmid,2,3 and Joseph Lau2

Patients at the same stage of a chronic disease may have had different rates of disease progression. The authors developed a mathematical modeling approach that allows reconstructing and comparing populations in terms of the disease progression rates of their participants when the disease onset and progression rates are unknown for individual patients. Human immunodeficiency virus 1 infection was used as an example. Both published and hypothetical models were used to describe the human immunodeficiency virus 1 epidemic (epidemic heterogeneity) and incubation and survival functions for different disease stages (individual heterogeneity). Reconstructions of populations with late disease (e.g., acquired immunodeficiency syndrome patients) show a marked predominance of rapid progressors, unless the incidence of new infections has been decreasing for a long time. Rapid progressors would also predominate in populations of acute seroconverters, unless diagnosis is based on repeated serologic screening rather than symptoms. Populations of patients who have not progressed beyond an early stage of the disease (e.g., patients with CD4 cell counts >500/μl) tend to overrepresent slow progressors, especially if the epidemic has been decreasing for a long time. With this approach, one can assess whether the target population of a clinical trial is comparable with other patient populations at different places and times. Epidemic and individual diversity may even affect trial results if patients with different progression rates experience different benefits from a treatment. By modeling the targeted populations in trials of early versus deferred antiretroviral treatment, the authors observed larger treatment benefits in trials in which rapid progressors probably predominated, compared with trials of slow progressors. Am J Epidemiol 1996;144:1074-85.

HIV; latency; models, statistical; randomized controlled trials; study design; survival

In epidemic diseases with chronic courses, patient populations classified as being at the same stage of the disease may not be similar. Heterogeneity may be caused by inherent biologic differences in the rates of disease progression among individual patients at the same disease stage; some patients may progress to a given stage faster than others. Furthermore, additional diversity may be caused by changes in the incidence of new cases of the disease in different places over time. As a result, patient populations at the same stage of the disease may have markedly different representations of slow and rapid progressors. When the onset of the disease is not recognizable and is followed by a variably long latent period, it is not possible to know whether a patient is a slow or a rapid progressor based only on his or her current disease stage.

A knowledge of the representation of rapid and slow progressors is important in comparing patient populations and in assessing whether the clinical experience obtained in one population can be directly extrapolated to other populations as well. In this report, we describe a mathematical modeling approach to assess the impact of epidemic and individual sources of heterogeneity on diversifying patient populations. The model can allow the evaluation of populations in terms of the rates of disease progression of their participants and is particularly useful when the time since the onset of the disease for individual patients and their rates of progression are unknown. Such mathematical modeling may help to assess whether the knowledge derived from specific clinical trials can be generalized and extrapolated to other patient populations at different geographic locations, at different times, and with dif-

Received for publication April 29, 1996, and accepted for publication June 18, 1996.

Abbreviations: AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus.

1 HIV Research Branch, Therapeutics Research Program, NIAID/NIH, Solar Building, Room 2C15, 6003 Executive Blvd., Bethesda, MD 20892. (Correspondence and reprint requests to Dr. Ioannidis at this address.)
2 Division of Clinical Care Research, New England Medical Center Hospitals, Boston, MA.
3 Biostatistics Research Center, New England Medical Center Hospitals, and Tufts University School of Medicine, Boston, MA.
Epidemic and Individual Population Diversity

Am J Epidemiol  Vol. 144, No. 11, 1996

different risk factors, which is an issue of considerable debate (1, 2).

As a practical example, modeling is applied to the patient populations targeted by trials of early versus deferred antiretroviral therapy in human immunodeficiency virus (HIV) type 1 (HIV-1) infection. HIV-1 has caused an epidemic disease with known extensive variability both in the rates of progression of individual patients and in the evolution of its epidemic locally and globally. In addition, the time of the primary HIV infection and rates of progression are unknown for the majority of the infected patients. Finally, we also explore the potential impact of these sources of heterogeneity on the results of clinical trials and their generalizability as exemplified by trials of HIV infection.

METHODS

General considerations and definitions

Consider an epidemic disease that progresses to its final outcome (e.g., death) through a late stage S with different rates of progression among individual patients. The time since infection, t, for a patient in stage S is an estimator of a patient’s rate of progression: Patients who are at this stage shortly after their infection are more rapid progressors than patients who have taken longer to be at the same stage after their infection.

Let us suppose that a clinical trial enrolls n patients at disease stage S. These n patients will be selected from a population pool of all patients at stage S. Let \( g_s(t) \) be the probability density of the incubation period, namely the first derivative of the incubation function \( G_s(t) \), which provides the probability that a patient will reach stage S within i months after being infected; and let \( I(t - i) \) be the survival function after stage S, representing the probability of not progressing to the final outcome for \( t - i \) months after reaching stage S. Then the probability that a patient is in stage S if the disease was contracted \( t \) months before the enrollment time is the sum of the products \( g_s(t) \times I(t - i) \), representing progression to stage S in \( t \) months followed by survival for \( t - i \) months, i.e.,

\[
U(t) = \int g_s(t) \times I(t - i) \, dt.
\]

Let \( T \) be the time before the enrollment time when the epidemic had its onset; and let \( X(T - t) \) be the epidemic function denoting the number of new infections at time \( t \) before the enrollment time. The number of patients that contracted the disease at a time \( t \) before the enrollment time and are currently at stage S is by definition equal to the product of \( X(T - t) \) and \( U(t) \). Therefore in the pool of patients with S, the proportion who contracted the disease \( t \) months ago would then be given by

\[
v(t) = U(t) \times X(T - t) / \int_0^t U(t) \times X(T - t) \, dt.
\]

The denominator of \( v(t) \) is the total number of patients at stage S when the trial enrolls patients. The \( n \) patients for the trial would be drawn from this pool. The numerator of \( v(t) \) is the number of patients at stage S who contracted the disease at a time \( t \) before trial enrollment. In the population pool, the cumulative proportion of patients with S that contracted the disease within \( t \) months would be given by the cumulative function \( V(t) \):

\[
V(t) = \int_0^t v(u) \, du.
\]

The population mean time from contracting the disease \( t_{\text{mean}} \) is

\[
t_{\text{mean}} = \int_0^t v(t) \, dt,
\]

and the population variance of \( t \), \( t_{\text{var}} \), is provided by

\[
t_{\text{var}} = \int_0^t (t - \text{mean})^2 \, v(t) \, dt - (\text{mean})^2.
\]

For large values of \( n \), there is a 95 percent chance that the estimated mean time in a sample of \( n \) patients enrolled in a clinical trial will satisfy:

\[
t_{\text{mean}} - 1.96 \times \sqrt{[t_{\text{var}}/n]} \leq t_{\text{mean}} \leq t_{\text{mean}} + 1.96 \times \sqrt{[t_{\text{var}}/n]}.
\]

The population median time since seroconversion can be derived by solving \( V(t) = 0.5 \) for \( t \).

If the epidemic is in steady state, i.e., \( X(T - t) \) is constant, then the proportion of patients in the pool of patients at stage S that contracted the disease \( t \) months ago would be given by

\[
v_{\text{steady}}(t) = U(t) / \int_0^t U(t) \, dt.
\]

It follows that the ratio of \( v(t) \) to \( v_{\text{steady}}(t) \) would be

\[
X(T - t) \times \int_0^t U(t) \, dt / \int_0^t U(t) \times X(T - t) \, dt,
\]

which means that at any value of time \( t \), the ratio of \( v(t) \) to \( v_{\text{steady}}(t) \) will be affected by \( X(T - t) \), that is by the course of the epidemic in relation to the time considered. Unless the incidence of new cases has been constant, \( v(t) \) will differ from \( v_{\text{steady}}(t) \).

It may also be interesting to consider clinical trials in which patients at disease stages earlier than S have been enrolled. First, let us consider patients who have not progressed to an initial early disease stage E. If \( G_E(t) \) is the incubation function for stage E, then the probability that a patient has been infected \( t \) months ago and has not progressed to stage E for at least \( t \) months would be \( F_E(t) = 1 - G_E(t) \). Among patients who have not progressed to stage E, the proportion
who contracted the disease $t$ months before the time considered would be

$$h(t) = [F_E(t) \times X(T - t)] \int_0^t F_E(u) \times X(T - t) \, dt.$$  

A second possibility is that diagnosis is made at the onset of the disease. Interventions would then aim at preventing the development of a late stage such as $S$ (secondary prevention) rather than the final outcome. If diagnosis at disease onset cannot be equally attained for all patients regardless of their subsequent rate of disease progression, but it depends on a feature that is differentially present in rapid and slow progressors, then the incubation function $G_{\text{onset}}(i)$ and its probability density function $g_{\text{onset}}(i)$ for the progression of patients diagnosed at the onset of the disease would be different from $G_S(i)$ and $g_S(i)$.

**Example: antiretroviral trials of HIV-1 infection**

This theoretical framework was applied to HIV-1 infection. For progression to late stage $S$, we considered progression to the acquired immunodeficiency syndrome (AIDS) with a clinical AIDS-defining illness (3). For progression to an early stage $E$, we considered progression to a CD4 count of 500/$\mu$L. Diagnosis at the onset of the disease was represented by diagnosis at the time of acute seroconversion. Mathematical models were used to describe the incubation, survival, and epidemic functions (4–7). We used published models, empirical models approximating cohort data, and hypothetical models to evaluate the diversity generated in the population pool of patients.

For the incubation function $G_S(i)$, we used (figure 1A) a Weibull model (5) with continuously increasing hazard for progression to AIDS and a median of 10 years (120 months). The introduction of therapy in the late 1980s has not allowed unbiased knowledge of the late natural history of the disease (4). Therefore, we also considered a log-logistic model with a smaller median (100 months) but with decreasing hazard of progression to AIDS in late years (figure 1A).

Survival after AIDS is generally short. For the survival function $l(t - i)$, we used the Weibull model $\exp[-0.015 \times t^{1.5}]$ that approximates published cohort data (8–10). It has a median of 12.8 months and predicts that one of six patients survives for 2 years after a clinical AIDS diagnosis.

In recent years, better care, therapy, and prophylaxis have halted disease progression and have continually improved survival after AIDS (8–11). Sensitivity analyses were based on the assumption that the function of the survival after AIDS depends on the time from the onset of the epidemic until the development of AIDS and that the incubation function to AIDS of HIV-infected patients depends on the time from the onset of the epidemic until the time infected. For both sensitivity analyses, empirical proportional hazards Weibull models were used for the nonstationary family of incubation and survival distributions with the time since the onset of the epidemic as a covariate of the hazard function. We did not use “time since infection models” (12, 13) because we feel these would be impractical to apply given the multitude of effective prophylactic and therapeutic measures that have been introduced episodically in the last decade (4, 13).

In the absence of systematic repeated serologic screening, very few patients are diagnosed at the time of their primary infection and seroconversion (14, 15). Patients with a manifest symptomatic seroconversion syndrome are more readily diagnosed than asymptomatic seroconverters, and it is known that symptomatic seroconverters are more likely to progress rapidly thereafter (14–16). We considered a Weibull and a quadratic model (figure 1B) for the incubation period to AIDS, $G_{\text{onset}}(i)$ of symptomatic acute seroconverters, consistent with published data from two cohort studies of such patients (17, 18) as compared with the log-logistic model used above for all HIV-infected patients (with symptomatic or asymptomatic seroconversion).

The function $F_E(i)$ for maintaining a CD4 count $>500/\mu$L is not precisely known, but cohort data suggest that the median CD4 count drops to 500/$\mu$L within about 4 years after the infection (17). We thus considered two plausible incubation functions based on log-logistic and Weibull models with a median time to progression of 4 years (figure 1C). The incubation function was considered to be stationary because no therapeutic interventions have been used widely at this early stage to affect the natural history of the disease.

For the epidemic function $X(T - t)$, we generally used a published back-calculation estimate of the course of the epidemic in the United States (6, 7), which assumes that the epidemic started in 1977 and peaked in 1985. Other polynomial functions were also used to describe hypothetical increasing and decreasing courses of the epidemic, and comparisons were made with steady-state conditions. All calculations and simulations were performed with Mathcad (19).

To also give practical examples from trials that have already been conducted, we considered published randomized, double-blind controlled trials of early versus deferred initiation of antiretroviral chemotherapy. A previous meta-analysis (20) was updated with MEDLINE searches until April 1996. The search strategy has been described in detail elsewhere (20). The protocol was expanded to also include studies of pri-
Epidemic and Individual Population Diversity

FIGURE 1. Incubation functions used in the calculations for different stages. A, probability of progression to acquired immunodeficiency syndrome (AIDS) among all human Immunodeficiency virus-infected patients according to the Weibull $1 - \exp \left( -0.0021 \times t^{2.616} \right)$ (5) and the log-logistic $1 - 1/(1 + \left(0.01 \times t^2\right))$ functions. B, probability of progression to AIDS for symptomatic acute seroconverters according to (a) the Weibull model $1 - \exp \left( -0.014 \times t^{1.1} \right)$ based on published data from a cohort of symptomatic acute seroconverters with 68% progression at 56 months (17), and (b) the quadratic model $t^2/10,000$ fitting data from a different cohort with 31% progression at 56 months and 100% progression in 100 months (18) as compared with (c) the log-logistic incubation function $1 - 1/(1 + \left(0.01 \times t^2\right))$ with median of 100 months (same as in 1A). C, probability of maintaining a CD4 cell count above 500/$\mu$L according to the Weibull model $1 - \exp \left( -0.01 \times t^{1.4} \right)$ and the log-logistic model $1 - 1/(1 + 0.01 \times t^{1.3})$, both with a median time of 4 years.

RESULTS

Populations of patients with late disease (clinical AIDS)

Stationary incubation and survival functions. In figure 2 are demonstrated three different scenarios for populations of patients with clinical AIDS (near the onset of an HIV epidemic, at the end of the rise of the epidemic, and as the epidemic is decreasing) as compared with when the epidemic has been in steady-state for a very long period. The figure shows that during the early phase of the epidemic or even shortly after the end of its rise, the patient population pool and samples from this pool are likely to have a marked predominance of very rapid progressors; for some patient subgroups the relative over- or underrepresentation in the enrolled population compared with steady-state conditions may be more than 10-fold. As the incidence of new infections decreases, the cumulative differences become less pronounced overall, although patients who progress to AIDS in average periods of time tend to be overrepresented with concomitant underrepresentation of both very rapid and very slow progressors.

In all cases, including the steady state, the popula-
tion of patients with clinical AIDS provides a markedly skewed picture of the population of HIV-infected people as a whole. Assuming Weibull incubation and survival functions, among all HIV-infected patients, the median time to be in AIDS is 11 years (132 months) after seroconversion. The respective median time from seroconversion is shorter in the population of patients with AIDS, regardless of the epidemic phase meaning that patients with late disease tend to be more rapid progressors than the whole population of HIV-infected patients. For example, the median time since infection for patients with AIDS at $T = 48, 110, \text{ or } 200$ months after the onset of the epidemic is 19, 46, and 106 months, respectively. If the epidemic had always been in steady-state, the respective median time would be 38, 82, and 123 months. The figures for the mean time are very similar. For example, the mean time since infection among patients with AIDS at $T = 110$ months after the onset of the epidemic is 45 months. The variance of $t$ is small, and there is a 95 percent probability that the mean time since infection among the participants of a trial of size $n = 100$ patients would be between 43 and 47 months. The calculations are not affected substantially if a log-logistic function is considered for the incubation distribution. For example, among the $n = 100$ patients with AIDS at $T = 110$ months after the onset of the epidemic, the mean time since infection would be expected to be 37 months with a 95 percent probability to lie between 36 and 39 months. These calculations imply that overall, rapid progressors are likely to be overrepresented in any trial of late disease.

Nonstationary incubation and survival functions. In table 1, the effects of changes in the incubation and survival functions of AIDS are shown over time. It can be seen that by assuming that either the median incu-
The effects on the population pool of patients with AIDS become more prominent as more time elapses from the onset of the epidemic; they are negligible for the first 5 years of the epidemic (not shown).

**Populations of patients diagnosed at disease onset (acute seroconversion syndrome)**

In figure 3, the probability distributions of the time for progressing to AIDS among diagnosed acute symptomatic seroconverters are provided. Similar to trials enrolling AIDS patients, trials assessing the efficacy of antiretroviral treatment at the time of symptomatic seroconversion are likely to enroll a surplus of rapid progressors, and the exact distribution of patients depends on the extent of the differential diagnosis of rapid progressors at the seroconversion phase. Although their exact population predictions differ substantially, both the Weibull and the quadratic model that we considered for the incubation period to AIDS for acute symptomatic seroconverters provide short estimates for the mean time to progression to AIDS (47 and 67 months, respectively). In a trial of \( n = 100 \) acute symptomatic seroconverters, there is a 95 percent probability that the mean time to progression to AIDS in the absence of treatment would be 39–55 months according to the Weibull model and 63–71 months according to the quadratic model. These estimates are shorter than any conservative figure for the mean incubation to AIDS in the population of all HIV-infected patients, which includes both symptomatic and asymptomatic seroconverters (3, 5).

---

**TABLE 1. Effect of changes in the incubation and survival functions of acquired immunodeficiency syndrome (AIDS) over time**

<table>
<thead>
<tr>
<th>Function changing over time*</th>
<th>Mean time since infection (years)</th>
<th>Estimated no. of AIDS patients†</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 years since the onset of the epidemic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>3.9</td>
<td>A</td>
</tr>
<tr>
<td>Incubation</td>
<td>3.9</td>
<td>0.78A</td>
</tr>
<tr>
<td>Survival</td>
<td>4.0</td>
<td>1.16A</td>
</tr>
<tr>
<td>Both</td>
<td>4.0</td>
<td>0.91A</td>
</tr>
<tr>
<td>18 years since the onset of the epidemic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>9.7</td>
<td>B</td>
</tr>
<tr>
<td>Incubation</td>
<td>10.0</td>
<td>0.91B</td>
</tr>
<tr>
<td>Survival</td>
<td>10.1</td>
<td>1.93B</td>
</tr>
<tr>
<td>Both</td>
<td>10.3</td>
<td>1.51B</td>
</tr>
</tbody>
</table>

*The incubation to AIDS is modeled with the Weibull function \( 1 - \exp\left[0.0021 \times e^{(0 - 0.05 \times \frac{t}{12})^{0.616}}\right] \), with \( k \) taking the values \(-0.05 \) and \( 0 \), depending on whether it is assumed that the incubation function is nonstationary (changing over time) or stationary. With the nonstationary assumption, the median time to patients infected in 1977, 12 years for patients infected in 1988, and projected to be 15 years for patients infected in 1997. The survival after AIDS is modeled with the Weibull function \( -0.024 \times e^{(0 - 0.1 \times (t - 1)^{0.5})} \), with \( k \) taking the values \(-0.1 \) or \( 0 \), depending on whether it is assumed that the survival function is nonstationary or stationary. With the assumption of nonstationary survival, the median survival after AIDS is 12 months for patients developing AIDS in 1981, 16 months in 1986 (9–10), and projected to be 36 months for patients developing AIDS in 1997. A backcalculation estimate (6, 7) is used for the epidemic function.

†The estimated number of patients with AIDS in the population compared with the number estimated under the assumption of stationary incubation and survival functions (A at 9 years and B at 18 years).
Populations of patients with early disease (CD4 count >500/μL)

Populations of patients that have not progressed beyond an early stage of the disease are likely to include a predominance of slow progressors. In the population of all HIV-infected patients, progression to CD4 < 500 may occur at any time according to rates determined by the incubation function. Among patients with CD4 > 500 at time \( t \) after infection, however, such progression may occur only at times larger than \( t \). As shown in figure 4B, the time since seroconversion that patients have not progressed to a CD4 count less than 500 can vary considerably depending on the course of the epidemic (figure 4A). This time is short when the incidence of new cases is rising and early in the epidemic (not shown). The effect of the course of the epidemic becomes more prominent if the epidemic is declining and especially if the incidence of new infections has been decreasing for a very long time. Under two such hypothetical situations occurring 25 years (300 months) or 33 years (400 months) after the onset of the epidemic (i.e., in 2002 or 2010, respectively), the population of patients with CD4 > 500 would be comprised mostly of very slow progressors. For example, assuming a log-logistic incubation function, we find that patients with more than 500 CD4 cells in 2010 would have avoided progression to CD4 < 500 for a median of 116 months (9.8 years) and a mean of 134 months (11.2 years). There is a 95 percent chance that the sample mean for \( n = 100 \) patients enrolled in a trial at that stage would be between 116 and 152 months. If a Weibull model is considered for the incubation function, the figures tend to be more modest. The mean and median are estimated to be 80 months (6.7 years), and there is a 95 percent chance that the mean for 100 enrolled patients would be between 68 and 92 months.

Practical examples from randomized trials of early versus deferred therapy

In table 2, the enrollment criteria and enrollment periods for the published trials of early versus placebo/deferred antiretroviral treatment (21–32) are shown. Generally, it can be seen that trials of patients with more advanced disease tended to be performed earlier in the epidemic. Time lags of 3–5 years (and even up to 8 years) from the onset of enrollment until the time of publication were typical in these trials. This means that the trial results were presented at a time when the epidemic could have been in a totally different phase than when the trials were initiated.

We analyzed in more detail according to our modeling approach five trials that have had a substantial clinical impact and have created controversies with their results. The other trials had relatively few death events (table 2). It should be remembered that all these trials were not designed with the concept of different individual rates of progression in mind, and epidemic and individual sources of heterogeneity were not anticipated in their design. Thus, mathematical modeling of their populations is used here as a hypothesis-generating tool.

The initial zidovudine study by Burroughs Wellcome (22), which showed the maximal benefit from zidovudine and established it as a cornerstone of HIV treatment, was conducted in early 1986 as the HIV epidemic was probably reaching a peak in the United States, and it enrolled patients with AIDS or severe AIDS-related complex. Its design resembles the situation shown in figure 1 for a study performed 110
Europe was in a phase equivalent to that of ACTG019 in the United States, i.e., with stable or declining probably performed at a time when the epidemic in Europe than in the United States (33), Concorde was patients with asymptomatic disease. Although we have less information for the course of the epidemic in

There has been only one randomized study (32) published to date on the treatment of primary HIV infection. This trial showed a large and significant incidence of new HIV infections (33). Concorde enrolled patients regardless of their CD4 cell counts provided they had not developed symptoms. This is a similar situation as the modeling of a population that has not developed a certain drop in the CD4 cell counts, i.e., has not progressed beyond an early stage of the disease. In that regard, Concorde probably also enrolled a preponderance of slow progressors and was probably also close to the situation depicted by tracing a in figure 4.

The VA298 study (22) of symptomatic patients with CD4 cell counts between 200 and 500/μl (26) showed no benefit from early zidovudine for those of the Burroughs Wellcome study and those of Concorde or ACTG019. There was a benefit from early therapy in terms of progression to AIDS, but no survival benefit. The study was performed at about the same time as ACTG019. A population with symptoms and CD4 < 500/μl is the exact opposite of the ACTG019 > 500 substudy population, and rapid progressors would thus predominate. However, VA298 excluded the population of patients with AIDS modeled in figure 1. Therefore, it probably had more rapid progressors than ACTG019 but lacked the very rapid progressors of the Burroughs Wellcome study.

There has been only one randomized study (32) published to date on the treatment of primary HIV infection. This trial showed a large and significant

- BW, Burroughs Wellcome protocol; AIDS, acquired immunodeficiency syndrome; ARC, AIDS-related complex; ACTG, AIDS Clinical Trials Group protocol; EACG, European-Australian Collaborative Group protocol; VA, Veterans Affairs Cooperative Study; ZDV, zidovudine; ACV, acyclovir. Data are derived from the original publications of the trials and exclude secondary publications with data from longer term follow-up of studies after unblinding.
- † Time from the onset of enrollment until the peer-reviewed publication of the trial results.
- ‡ Approximate data.
- § ACTG019 was composed of two substudies, one with patients of CD4 cell counts >500/μl (26) and one with CD4 cell counts of <500/μl (25).

months after the onset of the epidemic. Most likely, very rapid progressors were predominantly studied in this trial, which enrolled late stage patients early in the epidemic. It is possible that slow progressors were completely excluded inasmuch as patients should have been infected well before the mid-1970s to still qualify as slow progressors if they had AIDS by the mid-1980s. It is doubtful that the HIV epidemic started so early.

Later studies focused on patients with less advanced disease. For example, in the large ACTG019 trial, the subgroup with CD4 cell counts greater than 500/μl (26) showed no benefit from early zidovudine for asymptomatic patients with such high CD4 cell counts. Its design resembles the situation shown in figure 4 for a study performed 140 months after the onset of the epidemic, when the epidemic had started declining. There was probably a substantial preponderance of slow progressors in this trial.

The European Concorde trial (30) similarly found no benefit from early initiation of zidovudine among patients with asymptomatic disease. Although we have less information for the course of the epidemic in Europe than in the United States (33), Concorde was probably performed at a time when the epidemic in Europe was in a phase equivalent to that of ACTG019 in the United States, i.e., with stable or declining
benefit from antiretroviral therapy on acute seroconverters. Patients were mostly diagnosed and enrolled on the basis of symptoms. Therefore, this study probably enrolled a large proportion of rapid progressors, similar to the situations shown in figure 3 by tracings a and b.

Generally, these examples suggest that trials with a preponderance of rapid progressors tend to show large treatment benefits, whereas trials in which slow progressors probably predominated show no substantial benefit from the treatment. The results of the other trials with fewer events are also consistent with this trend. Future trials that consider differences in individual progression rates in their design are needed to further test and validate this finding.

**DISCUSSION**

In this paper, we have modeled and calculated how differences in individual rates of disease progression and in the course of the epidemic of a disease over time can diversify populations of patients who are classified as being at the same stage of the disease. Rapid progressors are overrepresented in populations of late disease stages, but their predominance is curtailed when the epidemic (the incidence of new infections) is decreasing for a long time. Slow progressors are overrepresented in populations of early disease stages, and their preponderance is enhanced when the epidemic is waning. Our mathematical modeling approach can allow a comparison of past, current, and future populations in terms of the rates of progression of their participants.

The representation of slow and rapid progressors may be different in populations involving patients at the same disease stage if these populations are studied at different times or at different places when and where the epidemic is in a different phase. Clinical trials sample patients under such diverse conditions. Moreover, time lags of 3 or more years from patient enrollment to the presentation of trial results commonly occur. Such time lags mean that the population on which the results of a trial were obtained may be substantially different from current patient populations even at the sites where the trial was performed. Similarly, multicenter trials and meta-analyses of trials are likely to engender substantial intercenter and intertrial heterogeneity unless the epidemic had a similar profile in all the participating centers and trials. Finally, the patient population targeted by a clinical study typically provides a skewed picture of all patients contracting the disease. This picture may be affected by changes in the diagnostic methodologies allowing improved and earlier diagnoses and by changes in the phases of the disease targeted for treatment.

Differences in the rates of disease progression of the study populations may also result in differences in the observed treatment effects if patients at high risk of disease progression do not experience the same benefit or harm from an experimental treatment as patients at low risk. There is evidence for some therapies that the magnitude of the treatment effect may depend on the baseline risk of the study patients (34). Examples include surgery for carotid artery disease (1), cholesterol-lowering regimens for cardiovascular disease mortality (35, 36), aspirin for the prevention of preeclampsia (37), and antiarrhythmics after myocardial infarction (38). It is possible that the baseline risk of progression may affect the efficacy of treatments for certain epidemic diseases as well, but past trials were not designed with this hypothesis in mind. Those who design future trials should try to take sources of epidemic and individual heterogeneity into account in their study design and in the interpretation of their results. This is important for identifying potential subgroups of patients with differential response to treatment.

Our analysis of trials on HIV-1 therapy exemplifies these issues. The natural history of HIV infection can be highly variable. Some patients develop AIDS and die within a few years after seroconversion (39) whereas others show no evidence of immune system dysfunction for at least 10 years after the primary infection (40). Moreover, the course of the HIV epidemic may be markedly different in New York City compared with Zaire or Thailand, and it may be markedly different in 1995 compared with 1985 or 2005. Marked differences exist even between neighborhoods in the same city (41). The epidemic may change even in the same neighborhood during the conduct of a clinical trial.

In a previous meta-analysis (20), we showed that there was a transient benefit from early initiation of antiretroviral therapy and that symptomatic patients experience a larger relative risk reduction than asymptomatic patients. It is conceivable that symptomatic patients enrolled in these trials may have been more rapidly progressing as a group relative to asymptomatic patients. As we showed with several examples from the modeling of trials of early versus deferred antiretroviral therapy, it is possible that differences among trials may be at least partially attributable to differences in the rates of disease progression of the participants.

Of course, heterogeneity in the results of clinical trials may stem from additional factors. For example, the duration of follow-up may be important when a therapy has time-limited efficacy (20). Possibly the epidemic or, in the case of HIV infection, the virus...
strains may overall become more (or less) virulent over time (42). Although hard to prove or disprove, there is no convincing evidence to support this hypothesis, and most recent investigations have failed to observe such a trend (43, 44). Finally, the introduction of newer and more effective therapies may affect the rates of disease progression of individuals and cohorts. The effect of therapy may be hard to assess with accuracy because some new therapies may be introduced episodically. In this paper, we also assessed the effect of nonstationary incubation and survival functions reflecting the introduction of effective therapeutic interventions. We have calculated how the lengthening of the incubation and survival periods tends to affect the pool of patients with late disease (e.g., AIDS) and how the overrepresentation of rapid progressors is exaggerated among these patients who have still progressed to a late disease stage despite the benefit of available therapies.

With the recent availability of more potent antiretroviral drugs and combinations, there is a current trend of shifting antiretroviral therapy to earlier stages (45). There is also a strong interest in improving early diagnosis and treating patients at the time of the primary infection (46). Our calculations suggest that future trials studying patients with early disease in communities where the epidemic has waned or in populations where the risk of infection has been decreasing for a long time will enroll a surplus of very slow progressors. Their results and their implications may be different compared with trials performed in communities where the epidemic is still rising or in certain groups in which the risk of infection is still high. Moreover, trials studying the effectiveness of treatment during the acute seroconversion phase may target different populations depending on how acutely infected individuals are diagnosed. If patients are identified on the basis of clinical symptoms, a preponderance of rapid progressors will probably be enrolled. Given the difficulty and rarity of making a diagnosis of primary infection (46), especially in the absence of symptoms, identification based on repeated serologic screening of high-risk individuals may also be proposed as it has been performed in certain high-risk cohorts (17). Such an approach, however, is very expensive and would identify a different population from a population diagnosed and enrolled on the basis of symptoms alone. It would be interesting to test prospectively whether the results of clinical trials using these two approaches are similar or not.

Given the existing population diversity in chronic diseases with wide epidemic and individual heterogeneity, it is important for clinical trials to define the stage of the disease and rates of progression of the enrolled patients as precisely as possible. In HIV infection, CD4 cell counts alone offer a satisfactory staging system for the advance of the disease but limited insight into whether a patient is a slow or a rapid progressor. Therefore, conventional subgroup analyses based on CD4 counts would not be able to adequately address the impact of the rates of disease progression on treatment outcomes. The use of other biologic markers with improved prognostic value, such as RNA viral load measurements (47–49) or CD38/lymphocyte activation antigen expression (50), may enhance the ability to design trials with better defined populations and assess whether specific groups of slow progressors experience different benefits from treatment than groups of rapid progressors. This would be an important step toward selectively targeting the patients who stand the greatest chance of deriving benefit from treatment. However, because it is unlikely that any biologic marker will ever be simultaneously perfect, convenient, and inexpensive, other factors, and especially the phase of the epidemic at a given time and place, may also have to be considered in designing future clinical trials and before making and generalizing treatment decisions based on their results.

The methodology presented in this report can be applied to diseases other than AIDS, provided one recognizes the heterogeneity among patients and a differential exposure to etiologic factors over time. Other infectious diseases such as viral hepatitis have extensive clinical variability among individuals and their incidence may vary over time. Neoplastic diseases also evolve differently in different patients, and specific population exposures to carcinogens can vary over time and place. In all situations, the implications of our approach are that patients who phenomenologically belong to the same disease stage may in fact be different and may also need to be treated differently for an optimal outcome. Differences in place, time, and the baseline risk of progression may decisively diversify the profile of seemingly similar populations and may influence the efficacy of and indications for specific treatments.

ACKNOWLEDGMENTS

Dr. Ioannidis was supported by an AIDS training grant from the National Institutes of Health (T32 AI07389; principal investigator, Dr. G. T. Keusch). This research was also supported in part by grant R01 HS 07782 from the Agency for Health Care Policy and Research of the US Public Health Service.
REFERENCES


