Progression to acquired immunodeficiency syndrome (AIDS) among persons infected with human immunodeficiency virus (HIV) varies considerably and may be influenced by factors such as age, smoking, number of male partners per year, and CD4 T-lymphocyte count. The loss of CD4 lymphocytes is known to be the dominant factor in the progression to AIDS. However, it is unclear whether the effect of the CD4 lymphocyte count is of such importance that persons with similar CD4 cell counts who have been infected for widely different lengths of time have the same risk of AIDS. While a CD4 count is easily obtainable, the precise amount of time since HIV infection is in most circumstances difficult to assess. In the present analysis, 259 Danish and 245 American homosexual men were followed for up to 14 years from 1981 to 1995. Two hundred and one persons seroconverted during the study period, and 112 had developed AIDS before the end of follow-up. CD4 lymphocyte count was highly correlated with the risk of developing AIDS ($p < 0.001$), but AIDS risk was not affected significantly by either age at infection, smoking, or number of male partners per year ($p > 0.20$ in all cases). Controlled for CD4 lymphocyte count, time since seroconversion was significant in explaining the risk of AIDS ($p = 0.018$), with a lower risk being seen during the first 3 years after seroconversion but no effect thereafter. These data confirm the central importance of CD4 lymphocyte level in the progression of HIV disease to AIDS, and suggest that rapid progression within 3 years of infection may be related to factors other than CD4 cell count. Am J Epidemiol 1997;145:629-35.

acquired immunodeficiency syndrome; CD4 lymphocyte count; cohort studies; HIV; homosexuality; risk factors

The incubation period between human immunodeficiency virus (HIV) infection and the development of acquired immunodeficiency syndrome (AIDS) is several years in most cases, but a continuous decline in immunity begins shortly after seroconversion. The great importance of the CD4 T-lymphocyte count and its central role in the immunopathology of AIDS was recognized early in the epidemic (1-3). One of the problems in analyzing the progression of HIV infection to AIDS is that it is not possible to observe the exact time of onset of HIV positivity. For each individual, one might typically know that the event occurred either before a certain date or between two specific time points. If knowledge of individuals' dates of seroconversion adds nothing to our ability to predict the onset of AIDS, this problem becomes irrelevant. Phillips et al. (4-6) have suggested that time from seroconversion is of little importance when the CD4 cell count is known. In a cohort of persons with hemophilia (4) and a cohort of Italian drug users (5), they found that the effect of time from seroconversion was insignificant when data were adjusted for CD4 lymphocyte count. This implies that persons infected long ago have the same risk of AIDS as those recently infected when their CD4 cell counts are the same. On the basis of these findings, Phillips et al. have described the application of Kaplan-Meier analysis to estimate the cumulative risk of AIDS as a function of declining CD4 cell count (6).

In this report, we investigate the prognostic effect of time from seroconversion to AIDS after controlling for CD4 lymphocyte count in HIV-positive homosexual men. We also analyze the effect of possible cofactors on the progression to AIDS, such as age at seroconversion, smoking status, and number of male partners...
per year. Age has been reported to affect progression to AIDS in some studies, such that older persons have faster progression than younger persons. An age effect has particularly been observed in study populations with a wide age range—e.g., persons with hemophilia (7). Previous analyses (8) have indicated that smoking increases CD4 lymphocyte count. Since the CD4 cell count is known to be associated with progression to AIDS, smoking might delay the onset of AIDS.

MATERIALS AND METHODS

Subjects

The study population consisted of 259 Danish and 245 American homosexual men. The Danish cohort was established in 1981 (9, 10), and follow-up was performed in 1982, 1983, 1984, 1987, 1989, and 1992. Participants donated blood for serum samples and completed a self-administered questionnaire on lifestyle, sexual practices, smoking status, and exposure to drugs. The serum samples were screened for HIV antibodies, and, in all years except 1981, the CD4 cell counts of participants were determined. The US cohort was established in 1982 and consisted of 160 homosexual men from Washington, DC, and 85 homosexual men from Manhattan, New York City (11, 12). Examinations took place every year from 1982 to 1992. The questionnaire and diagnostic criteria used were the same as those in Denmark. In Denmark and the United States, data on the cohorts were updated with respect to AIDS and vital status on January 1, 1995, and January 1, 1994, respectively.

The age profile was the same in the two cohorts. At enrollment, the participants ranged in age from 17 years to 73 years (median, 33 years), but 70 percent were between ages 25 and 40 years. In both cohorts, the number of male partners per year was very high in the early 1980s. In 1982 and 1983, 61 percent of the men in Denmark and 68 percent of the US men had more than 10 partners per year, while only 20 percent of the men had more than 10 partners per year. Therefore, the number of male partners per year was treated as a time-dependent covariate. In the United States, 60 percent of the men were nonsmokers at their first examination, whereas in Denmark this figure was only 40 percent.

Twenty-one Danes were HIV-positive at the first examination, and by 1994 another 48 were known to have become HIV-infected. AIDS had developed in 37 Danish men by the closure of the study. In the US cohort, there were 84 HIV-prevalent cases, and another 49 persons seroconverted between 1982 and 1994. In total, 76 US men developed AIDS. However, one prevalent case from New York had AIDS at entry into the study and was not included in the analyses.

Estimation of date of seroconversion

For the 104 prevalent cases, only the date of their first positive HIV test was available. Since the disease is assumed to have appeared slightly later in Washington, DC, the date of seroconversion for these persons was estimated as the midpoint between January 1, 1979 (New York City) or January 1, 1980 (Washington, DC) and the date on which they were first seen to be HIV-positive. Thus, persons from New York City who were HIV-seropositive at their initial visit in 1982 were assumed to have seroconverted in July/August 1980, while persons in Washington, DC, were presumed to have seroconverted approximately 7 months later.

In estimating dates of seroconversion for the incident cases, we used a method which addressed the interval censoring problem. This methodology is based on a generalized linear model with a logarithmic link function developed by Becker and Melbye (13) for cases in which the subjects are examined at equal time points, and has been further developed by Carstensen (14) to cover situations where the persons are not examined at equal time points. In short, this results in a piecewise exponential “survival function” for the time of seroconversion, and the date of seroconversion is estimated as the median “survival time” in the interval wherein the person is known to have seroconverted.

Adjustment of the CD4 lymphocyte count

During the years of the study, there was close cooperation between the research teams in Denmark and the United States to facilitate comparison of results from the two countries. Prior to 1987, HIV serology and CD4 measurements were evaluated in the same laboratory using similar methodology. Most recently, these analyses have been standardized and are now undertaken in each individual country. However, it is possible that the change in methodology over time might have influenced the variability of the CD4 cell

count. Furthermore, the individual variability of CD4 counts can be substantial and may be influenced by initial clinical status. Studying the HIV-negative samples gave us an estimate of the magnitude of variation. Analysis of variance on log-transformed CD4 measures from HIV-negative samples showed significant effects of both subject and time of measurement in both countries \((p = 0.0001)\), and this led to adjustment of the measured CD4 counts. Estimating the effect, \(\theta_{ij} = 1982, \ldots, 1992\), of calendar time (time of measurement), based on the CD4 counts for the HIV-negative persons, all CD4 counts were adjusted by this factor:

\[
\text{Adjusted } CD_4 = \theta_j CD_4_j.
\]

**Follow-up**

Since the main purpose of this investigation was to study the influence of certain risk factors on AIDS risk in HIV-positive homosexual men, we assumed in the further analyses that the time of seroconversion was known. We then considered only the 201 HIV-positive homosexual men and information collected from the estimated date of seroconversion to the development of AIDS or loss to follow-up. Each citizen in Denmark is assigned a unique 10-digit identification number, which permits accurate linkage of information from different registries. Information on AIDS and vital status was obtained via the national AIDS registry (mandatory registry) and the Central Person Registry. The Central Person Registry keeps updated files on all residents of Denmark and documents such demographic variables as death and migration. Follow-up of the US cohort was conducted annually with interview reports of events by subjects or their doctors. AIDS was defined as clinically manifest AIDS; thus, a diagnosis based solely on a low CD4 cell count (using the 1993 definition) was not considered AIDS-defining in our analysis. In Denmark and the United States, four and three persons, respectively, died of causes other than AIDS. These persons were censored at the time of death.

**Statistical methods**

To investigate which factors might have an effect on progression to AIDS, we used the Cox proportional hazards model with time-dependent covariates \((15)\). The basic time scale used was time since estimated date of seroconversion. Prevalent cases were observed with delayed entry (left-truncated), and therefore their “at-risk” period started only at the time at which they were first seen to be HIV-positive. In the analyses including CD4 cell counts, the individual “at-risk” period started at the time of the first CD4 lymphocyte measurement after the first positive HIV test. The covariates were analyzed separately because of the relatively few AIDS cases. Age at seroconversion was analyzed as both a continuous and a dichotomous variable: age minus the mean age, and age above or below the mean age. Number of male partners per year was defined as a dichotomous variable depending on whether the number of partners was above five or less than or equal to five. CD4 cell count was treated as a continuous (linear) variable, a log- or square root-transformed variable, and a grouped variable. CD4 count was divided into five groups: 0–100, 100–200, 200–500, 500–800, and >800 cells/mm³. CD4 count, smoking status, and number of male partners per year were all fitted as time-dependent covariates; i.e., the individual values were continually updated throughout the follow-up period.

The proportional hazards assumption was investigated using log-log survival plots, and results seemed satisfactory, whereas neither a logarithmic transformation nor a square root transformation of the CD4 count data satisfied the assumption of a log-linear effect.

Using calendar time as the basic time scale in the Cox regression model, as suggested by Phillips et al. \((5)\), it is possible to investigate the effect of time from seroconversion as well. However, this approach was not suitable for the present data, since most individuals were infected during the early part of the study, making time from seroconversion and calendar time highly correlated. Therefore, for analysis of the effect of time from seroconversion, we used a Poisson regression model \((15)\). The AIDS rate was assumed to be piecewise constant in different groups of CD4 counts and different time intervals since seroconversion. Again, the CD4 count was continually updated throughout the follow-up period and then data were grouped into the five categories described above. Time since seroconversion was divided into number of years from seroconversion (range, 3–9 years), and furthermore data were grouped by country. The groups were dynamic in the sense that not only did a person change time group during follow-up but he might change CD4 group as well, when a new CD4 count was made.

Although Cox regression analysis and Poisson regression analysis model the hazard of AIDS and the rate of AIDS, respectively, we will, for convenience, refer to the estimates as relative risks.

**RESULTS**

One hundred and twelve of the 201 HIV-positive subjects had developed AIDS before the end of the study. Because seven persons died of other causes before AIDS was diagnosed, AIDS survival time was
regarded as progression time to AIDS provided that the person did not die from other causes. A log-rank test for the effect of country was borderline significant ($p = 0.08$), and there was a tendency for the Danish men to have a longer time to progression than the US men. Therefore, in further analyses we stratified by country. When we fitted a Cox model stratified by country, no significant effect of smoking status, drug use, age at seroconversion, or number of male partners per year was found. However, there was a clear effect of CD4 lymphocyte count. The relative risk estimates are shown in table 1. Analyzing CD4 count as an untransformed time-dependent covariate, we found that the overall risk of AIDS increased by almost 60 percent for each CD4 count decline of 100. Investigation of the grouped CD4 count data showed no significant increase in the risk of AIDS when the CD4 count dropped from >800 to 500–800 ($p = 0.72$) or 200–500 ($p = 0.13$). However, a CD4 count between 100 and 200 increased the risk more than nine times, and a CD4 count less than 100 increased the risk 30 times relative to the risk for a person with a CD4 count greater than 800. We investigated the effect of the CD4 count measured at the last negative test and the first positive test, dividing the measured CD4 counts into two groups. To make the groups approximately equally sized, values for CD4 count at the last negative test were divided at 700 cells/mm$^3$ and values for CD4 count at the first positive test were divided at 550 cells/mm$^3$. Time to AIDS was not affected by a high or low CD4 count prior to seroconversion. However, if the first postseroconversion count was less than or equal to 550, the risk of AIDS doubled. Results were similar when the unadjusted CD4 counts were analyzed.

Fitting the Poisson regression model, a test for interaction between time from seroconversion and CD4 count showed that there was no interaction ($p = 0.40$). This means that the effect of the CD4 count on the progression to AIDS is the same at any time. Interactions between country and the other two covariates (time from seroconversion and CD4 count) were also investigated, but no interactions were found. (Because these tests had little power to detect possible interactions, they were performed in a model with only half the number of groups.) Table 2 shows the numbers of person-years at risk and AIDS cases according to CD4 count and time from seroconversion.

In the model without interaction terms, we found that both CD4 count and time from seroconversion were significant in explaining the risk of AIDS ($p = 0.018$ for time, $p < 0.001$ for CD4 count), but there was no effect of country ($p = 0.58$). There was a clear trend in the effect of the CD4 count such that the risk increased with decreasing CD4 count. Table 3 shows the estimated relative risks of AIDS. The estimated risks in the different CD4 groups are to be interpreted as the risk in each group relative to the risk associated with a CD4 count greater than 800 for CD4 counts observed at the same time from seroconversion. Similarly, the expected relative risk in each time interval represents the risk relative to that seen 5–6 years after seroconversion at the same CD4 cell count. The table indicates that the risk of AIDS was four times lower than

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Relative risk estimate</th>
<th>95% confidence interval</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of male sex partners per year (&gt;5 vs. ≤5)</td>
<td>1.04</td>
<td>0.70–1.55</td>
<td>0.84</td>
</tr>
<tr>
<td>Smoking (yes/no)</td>
<td>1.26</td>
<td>0.84–1.88</td>
<td>0.27</td>
</tr>
<tr>
<td>Age at seroconversion</td>
<td>1.08</td>
<td>0.95–1.23</td>
<td>0.47</td>
</tr>
<tr>
<td>Age greater than mean age at seroconversion</td>
<td>1.27</td>
<td>0.87–1.66</td>
<td>0.21</td>
</tr>
<tr>
<td>CD4 lymphocyte count§ (cells/mm$^3$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;800</td>
<td>1.59</td>
<td>1.39–1.82</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>500–800</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;200–500</td>
<td>3.05</td>
<td>0.26–6.33</td>
<td>0.72</td>
</tr>
<tr>
<td>&gt;100–&lt;200</td>
<td>9.92</td>
<td>2.27–43.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>≤100</td>
<td>30.66</td>
<td>6.89–136.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CD4 cell count &lt;700 at last negative test</td>
<td>1.26</td>
<td>0.50–2.66</td>
<td>0.54</td>
</tr>
<tr>
<td>CD4 cell count ≤550 at first positive test</td>
<td>2.37</td>
<td>1.40–4.01</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

* AIDS, acquired Immunodeficiency syndrome; HIV, human Immunodeficiency virus.
† Single factor analyses were performed. The variables (except age) were fitted as time-dependent covariates in a Cox model stratified by country (Denmark vs. the United States).
‡ Relative risk per 5-year increase.
§ Relative risk per 100 cells/mm$^3$ decrease.
within 3 years than it was 5–6 years from seroconversion, while the risk was steady after 3 years \((p = 0.80)\).

Figure 1 shows the estimated AIDS risks by time from seroconversion relative to AIDS risk 5–6 years from seroconversion without controlling for the CD4 count (dotted line), as compared with the relative risks presented in table 3 (full line). The figure shows that the risks after 8 years were overestimated when we did not control for the CD4 count. This is probably explained by the fact that after 8 years, the CD4 count has become very low in most cases—i.e., when controlling for the CD4 count, the effect of time from seroconversion is reduced.

Primary prophylactic treatment may delay the onset of AIDS. However, since treatment approaches are complex, including many regimens and variable lengths of prophylaxis, treatment was not considered directly as a cofactor in the analysis. Treatment became generally available by January 1, 1988, but we found no evidence of a decreased risk of AIDS after this date.

The effect of time from seroconversion was only seen within the first 3 years and could have been due to the large number of prevalent cases. When we analyzed the prevalent and "incident" cases separately, the findings were almost similar. When we controlled for the CD4 count, however, the effect of time from seroconversion was reduced.

seroconversion was only borderline significant \( (p = 0.048) \) among the “incident” cases. (The tests were performed in a model with only half the regular number of groups.)

The robustness of the estimated date of seroconversion was investigated using, respectively, the earliest possible and latest possible dates of seroconversion instead. However, the basic conclusions remained the same.

As noted above, there was no significant difference between the effects of time in the two countries. However, analyzing the two countries separately, we found that the effect of time was insignificant in the smaller Danish cohort (although with a trend in the same direction) but significant in the US cohort.

Assuming that the CD4 count can only be considered valid for 1 year after the date of measurement, an analysis was performed censoring subject follow-up at 1 year after the most recent CD4 count. This censoring pattern reduced the number of “observed” AIDS cases somewhat, but gave results similar to those described above.

**DISCUSSION**

Our results suggest that none of the covariates studied here (smoking status, drug use, number of male partners per year, and age at seroconversion) had any additional value over that of CD4 count in assessing the risk of AIDS. Other studies, particularly those conducted among hemophiliacs, have shown that age at seroconversion is especially correlated with progression to AIDS (7). The insignificant result found in this study (although with a trend in the same direction) might have been caused by the relatively small variation in age between subjects, 70 percent of whom were between ages 25 and 40 years at HIV seroconversion. There was a tendency for the Danish men to have a slower progression to AIDS than the US men. We therefore kept country as a cofactor in our analyses, even though there was no significant difference between the two countries. It is possible that the socialized health care system in Denmark, which offers the same high standard of medical care to all citizens free of charge, may have resulted in more extended prophylactic treatment in the Danish men after they were infected with HIV. This could have postponed their time of AIDS diagnosis. Because of the follow-up procedure for the US cohort, it is also possible that HIV-positive men who dropped out of the cohort and as such were censored in our analyses had a better outcome than those who remained in the cohort.

The analysis showed that the CD4 lymphocyte count is highly associated with progression to AIDS, and this corresponds well with previous findings (1–3, 11). The results were not based on any assumptions concerning the pattern of CD4 lymphocyte loss, but we adjusted the data for a possible calendar effect due to changes in measurement techniques. We found no association between CD4 count prior to HIV infection and the risk of AIDS, but definitive conclusions on this point would require a larger data set concerning measurement prior to HIV infection.

Primary prophylactic treatment may delay the onset of AIDS and may have influenced these findings. However, we found no decrease in the risk of AIDS after January 1, 1988, when prophylactic treatment became generally available. This is probably due to a strong correlation between calendar time and time from seroconversion in closed cohorts such as these.

In the Poisson regression analysis, we found that the CD4 count, as expected, was related to the risk of AIDS. Time from seroconversion was an important factor as well, even when data were controlled for CD4 count. The effect of time was only seen within the first 3 years, and the result should therefore be interpreted with some caution because of the relatively large fraction of prevalent cases in the two cohorts analyzed. However, analyzing the prevalent cases separately gave us results similar to those from the original model. Furthermore, using an estimated date of seroconversion as the earliest or latest possible date did not change the basic findings. In these subjects, CD4 count was measured, at most, every year. Thus, since we updated the CD4 count continually throughout the follow-up period, the CD4 groups to which we assigned people could in some cases have been quite unrealistic. However, censoring subject follow-up 1 year after the most recent CD4 count did not change the conclusions.

Our results differ from those reported by Phillips et al. (4, 5). However, the cohorts we studied were substantially different in terms of HIV exposure route (sexual activity vs. exposure to blood products or sharing of injecting equipment) and other aspects of lifestyle and medical care. In addition, we studied events over a 14-year time span from seroconversion. In the study by Phillips et al. (4), time zero was taken as August 1, 1985, and time from HIV seroconversion to AIDS then ranged from 0 to 6 years, while in our data, it ranged from 0 to 14 years.

One can obtain enhanced effects when the measurement error of a numerical explanatory variable in a regression model is correlated with the variable being examined (16). In our data, the measurement error of the CD4 counts may have been correlated with the true CD4 values (and thereby time), since low counts are likely to have smaller variance than high counts. To
minimize any such exaggeration, we treated the CD4 count in our model as a categorical covariate.

We conclude that a low CD4 count is the most powerful predictor of AIDS risk, but knowledge that a person has only recently seroconverted (within 3 years) adds additional information—namely, that he is at lower risk of AIDS than someone with a longer duration of HIV infection and the same CD4 count. In the present data, a small fraction of approximately 12 individuals had one or two transiently low CD4 counts close to seroconversion. Possibly, the strain of HIV and the host’s response to initial infection cause aberrations in the CD4 count that, in the early years of infection, are not tightly linked to functional cell-mediated immunity. These topics may provide fruitful avenues for further research.

ACKNOWLEDGMENTS

This study was supported by grants from the Danish Medical Research Council (nos. 1993-69-000-8 and 12-1617-1), the Danish National Research Foundation, and the US National Cancer Institute (contract no. 1-CP-61013). The activities of the Danish Epidemiology Science Centre are financed by a grant from the Danish National Research Foundation.

REFERENCES