TRANSACTIONS OF SOCIETY OF ACTUARIES
1988 VOL. 40 PT 1

AIDS: SURVIVAL ANALYSIS OF PERSONS TESTING HIV+

HARRY H. PANJER

ABSTRACT

The purpose of this paper is to provide a survival analysis of persons in the various stages of HIV infection typically leading to AIDS and ultimately death. The model used is a continuous time Markov process with a constant intensity for each stage. It is shown that this model adequately describes the data which originated from a German longitudinal study. The data were previously analyzed using less formal methods in the comprehensive paper of Cowell and Hoskins dealing with the effect of HIV infection on life insurance. This paper should be of special interest to health insurers since it deals with distribution of duration in each stage of progression of the disease.

INTRODUCTION

The progression of persons infected with the Human Immunodeficiency Virus (HIV) to Acquired Immune Deficiency Syndrome (AIDS) is being studied at the Centre for Internal Medicine of the University of Frankfurt [1]. The longitudinal study follows subjects in groups at high risk of AIDS through various stages from good health with an HIV+ status to death primarily caused by AIDS.

In the study 543 subjects were observed during the study period from 1982 through 1985; 377 were HIV+ at the time they were initially observed; 307 were observed for at least three months of which 259 were HIV+ upon initial observation.

The Walter Reed Staging Method [6] was used as a basis for classifying subjects:

1a	(At-Risk)	Healthy persons at risk for HIV infection, but testing negative;
1b	(HIV+)	Otherwise asymptomatic persons testing HIV+;
2a	(LAS)	Persons with HIV infection and lymphadenopathy syndrome (LAS), together with moderate cellular immune deficiency;
2b	(ARC)	Patients with HIV infection and LAS, together with severe cellular immune deficiency (AIDS-Related Complex, or ARC);
3	(AIDS)	Patients with AIDS.

The sixth stage was death.

Table 1A classifies the number of persons observed during the study by the length of time under observation and the initial stage; that is, the stage at the time the subject was first observed in the study. Table 1B gives the corresponding number of persons whose condition worsened at least one stage during the study period. Table 1C gives the ratio of Table 1A to Table 1B as a percentage. These tables are taken directly from the paper by Cowell and Hoskins [2, Part 2, p.19].

TABLE 1

Range of Observation Periods	Stage 1a (At-Risk)	Stage 1b (HIV+)	Stage 2a (LAS)	Stage 2b (ARC)	Stage 3 (AIDS)	All Stages
A) Number of Patients Observed by Stage and Observation Period						
3-6 months	10	9	21	8	6	54
6-12 months	14	18	51	29	9	121
12-24 months	21	20	29	20	7	97
24-36 months	3	5	19	7	1*	35
All Periods	48	52	120	64	23	307
B) Number of Patients Observed Whose Health Worsened by at Least One Stage, or Who Died during the Observation Period						
3-6 months	1	1	3	0	4	9
6-12 months	6	10	20	3	6	45
12-24 months	9	15	14	10	5	53
24-36 months	2	4	14	4	0*	24
All Periods	18	30	51	17	15	131
C) Percentage of Patients Observed Whose Health Worsened by at Least One Stage, or Who Died during the Observation Period						
3-6 months	10%	11%	14%	0%	67%	17%
6-12 months	43%	56%	39%	10%	67%	37%
12-24 months	43%	75%	48%	50%	71%	55%
24-36 months	67%	80%	74%	57%	0%*	69%
All Periods	38%	58%	42%	27%	65%	43%

*One patient with AIDS was still alive 28 months after diagnosis of Kaposi's sarcoma; all others with AIDS had died before the end of 24 months.

It should be noted that only information regarding the initial stage (that is, the stage at the time an individual entered the study) is available from these tables. An individual passing through several stages during the period of observation is indistinguishable from an otherwise identical person but moving only to the next stage.

The later sections of this paper deal with a formal analysis of these data for the purpose of making inference about the distribution of lifetime for persons in the various stages.

THE MODEL

As in Cowell and Hoskins [2, Part 2, pp. 4,5], it is assumed that individuals progress through the successive states in order and do not return to a previous state. In medical terms, this means that a person's condition (according to the Walter Reed Staging Method) can remain the same or deteriorate but can never improve.*

We model progression from stage to stage by assuming that a person in a given stage of HIV infection is subject to an intensity function (force of progression, hazard rate) that depends only upon the stage and not upon other factors such as age, sex and the length of time in the stage.

Let $\mu_j, j = 1a, 1b, 2a, 2b, 3$, denote the intensity function of progression from stage j to the next stage. Let T_j denote the time in stage j. Then, the probability that a person just entered stage j will remain in stage j for at least t years is

$$Pr\{T_j > t\} = \exp\{-\int_0^t \mu_j dt\} = e^{-t\mu_j}; \tag{1}$$

the cumulative distribution function (cdf) of T_j is

$$F_{T_j}(t) = \Pr\{T_j \leq t\} = 1 - e^{-t\mu_j}; \tag{2}$$

and the probability density function (pdf) of T_j is

$$f_{T_j}(t) = \mu_j e^{-t\mu_j}. \tag{3}$$

This is the exponential distribution with mean $1/\mu_j$ and variance $1/\mu_j^2$.

A consequence of this model is that the times $T_j, j = 1a, 1b, 2a, 2b, 3$, are stochastically independent. Furthermore, the memoryless property of the exponential distribution means that the length of time that a person has been in the current stage is irrelevant for our purposes and that the expected time of progression to the next stage is the same for all persons in the stage; that is, it is independent of the time already in the current stage.

is, it is independent of the time already in the current stage.

can be easily computed. For example, the random variable denoting the time from progression to AIDS (stage 3) from a positive HIV test (stage 1b) is

$$T_{1b} + T_{2a} + T_{2b} \tag{4}$$

*Seven of 307 subjects observed for more than three months improved at least one stage. This is attributed to the possibility of misdiagnosis of one of the stages, an event that can be expected to occur since some judgment is involved [1, p. 1178].

with mean

$$\mathrm{E}[T_{1b}] + \mathrm{E}[T_{2a}] + \mathrm{E}[T_{2b}] = \frac{1}{\mu_{1b}} + \frac{1}{\mu_{2a}} + \frac{1}{\mu_{2b}} \tag{5}$$

and variance

$$\mathrm{Var}[T_{1b}] + \mathrm{Var}\,[T_{2a}] + \mathrm{Var}[T_{2b}] = \frac{1}{\mu_{1b}^2} + \frac{1}{\mu_{2a}^2} + \frac{1}{\mu_{2b}^2} \tag{6}$$

due to independence of the T_j's.

The exact distribution function of $T_{1b} + T_{2a} + T_{2b}$ is easily obtained by integration. We first obtain the pdf for a pair of exponential random variables.

$$f_{T_1+T_2}(t) = f_{T_1} * f_{T_2}(t) = \int_0^t \mu_1 e^{-\mu_1 s} \mu_2 e^{-\mu_2(t-s)}\, ds$$

$$= \frac{\mu_2}{\mu_2 - \mu_1} f_{T_1}(t) + \frac{\mu_1}{\mu_1 - \mu_2} f_{T_2}(t) \tag{7}$$

where * indicates the convolution operator. Consequently,

$$f_{T_1 + T_2 + T_3}(t) = f_{T_1} * f_{T_2} * f_{T_3}(t)$$

$$f_{T_1+T_2+T_3}(t) = \frac{\mu_2\mu_3}{(\mu_2-\mu_1)(\mu_3-\mu_1)} f_{T_1}(t) + \frac{\mu_1\mu_3}{(\mu_1 - \mu_2)(\mu_3-\mu_2)} f_{T_2}(t)$$

$$+ \frac{\mu_1\mu_2}{(\mu_1-\mu_3)(\mu_2-\mu_3)} f_{T_3}(t). \tag{8}$$

In general, it can be shown [4, p. 79] that the distribution of $T_1 + T_2 + \ldots + T_n$ is

$$f_{T_1+T_2+\ldots+T_n}(t) = \sum_{i=1}^{n} \left\{ \prod_{j \neq i} \left(\frac{\mu_j}{\mu_j - \mu_i} \right) \right\} f_{T_i}(t). \tag{9}$$

Similarly, it can be shown that the probability that an individual in any stage (arbitrarily labeled 1) will pass through stages 2,3,. . ., $n-1$ and be in stage n exactly t years later is

$$\sum_{i=1}^{n} \left\{ \prod_{j=1}^{n-1} \mu_j \right\} \left\{ \prod_{j \neq i} \frac{1}{\mu_j - \mu_i} \right\} e^{-\mu_i t}. \tag{10}$$

This allows for exact evaluation of probabilities to answer questions such as "What is the probability that an HIV+ individual will develop AIDS within three years?" or "Of 1,000 HIV+ persons, how many can we expect to have AIDS or have died within five years?"

In the next two sections the parameters of the model are estimated and the model is tested for validity against the observed data given in Table 1.

ESTIMATION OF MODEL PARAMETERS

Consider a single stage with constant intensity μ. Since the times in each stage are independent, we consider the stages separately and drop the subscript j for notational convenience.

Notation:

μ = "force of progression" to next stage,
d_i = number of persons progressing for those in observation period i, $i = 1, 2, 3, 4$,
n_i = number of persons observed in observation period i, $i = 1, 2, 3, 4$,
p_i = probability of not progressing if in observation period i, $i = 1, 2, 3, 4$,
$q_i = 1 - p_i$ = probability of progressing at least one stage.

The likelihood function is

$$L(\mu) = \prod_{i=1}^{4} \binom{n_i}{d_i} (1 - p_i)^{d_i} p_i^{n_i - d_i}. \tag{11}$$

The loglikelihood function is

$$\mathcal{L}(\mu) = \sum_{i=1}^{4} \left\{ \log \binom{n_i}{d_i} + d_i \log(1 - p_i) + (n_i - d_i) \log p_i \right\}. \tag{12}$$

Note that only p_i is a function of μ and will be specified later.

The maximum likelihood estimator (MLE) $\hat{\mu}$ of the parameter μ is obtained as the maximum of $L(\mu)$ or equivalently of $\mathcal{L}(\mu)$.

Differentiating the loglikelihood yields

$$\frac{\partial \mathcal{L}(\mu)}{\partial \mu} = \sum_{i=1}^{4} \frac{d_i}{1 - p_i} \left(-\frac{\partial p_i}{\partial \mu} \right) + \frac{n_i - d_i}{p_i} \frac{\partial p_i}{\partial \mu}$$

$$= \sum_{i=1}^{4} -\frac{\partial p_i}{\partial \mu} \left\{ \frac{d_i}{1 - p_i} - \frac{n_i - d_i}{p_i} \right\}. \tag{13}$$

Solving

$$\frac{\partial \mathcal{L}(\mu)}{\partial \mu} = 0 \tag{14}$$

should yield the MLE $\hat{\mu}$.

We use a Newton-Raphson method to solve equation (14) for μ. Let

$$f(\mu) = \frac{\partial \mathcal{L}(\mu)}{\partial \mu} \tag{15}$$

$$f'(\mu) = \frac{\partial f(\mu)}{\partial \mu}. \tag{16}$$

Then

$$f'(\mu) = -\sum_{i=1}^{4} \left\{ \frac{\partial^2 p_i}{\partial \mu^2} \left[\frac{d_i}{1 - p_i} - \frac{n_i - d_i}{p_i} \right] + \left(\frac{\partial p_i}{\partial \mu^2} \right)^2 \left[\frac{d_i}{(1 - p_i)^2} + \frac{n_i - d_i}{p_i^2} \right] \right\}. \tag{17}$$

Beginning with an initial estimate of $\hat{\mu}_0$ the successive estimates are obtained as

$$\hat{\mu}_{n+1} = \hat{\mu}_n - \frac{f(\hat{\mu}_n)}{f'(\hat{\mu}_n)}. \tag{18}$$

Now we make some assumptions regarding the exact time-on-study for the subjects. This is necessary since, for example, for a given individual observed for between 24 and 36 months, we require the exact time of observation. First, we assume that each subject was observed up to the midpoint of the interval, for example, 30 months for each subject in the 24–36 month interval. Then we repeat the exercise using a random censoring mechanism.

A. Assume exposure to the midpoint of the interval, then we obtain the probabilities

$$p_1 = \int_{3/8}^{\infty} \mu e^{-\mu t}\, dt = e^{-(3/8)\,\mu}$$

$$p_2 = \int_{3/4}^{\infty} \mu e^{-\mu t}\, dt = e^{-(3/4)\,\mu}$$

$$p_3 = \int_{3/2}^{\infty} \mu e^{-\mu t}\, dt = e^{-(3/2)\mu}$$
$$p_4 = \int_{5/2}^{\infty} \mu e^{-\mu t}\, dt = e^{-(5/2)\mu}. \tag{19}$$

Then

$$\frac{\partial p_i}{\partial \mu} = \frac{\partial}{\partial \mu} e^{-\mu t_i^*} = -t_i^* e^{-\mu t_i^*} \tag{20}$$

and

$$\frac{\partial^2 p_i}{\partial \mu^2} = t_i^{*2} e^{-\mu t_i^*} \tag{21}$$

where t_i^* is the midpoint of interval i.

B. Assume a uniform random censoring mechanism for each observation. Let t denote the exact time of observation. If the limits of the right-hand end of observation period i are a_i and b_i, we have $a_i < t < b_i$. Since we have no information about the relative likelihood of the possible censoring times t, we treat t as a random variable with a uniform (a_i, b_i) distribution making all censoring times t equally likely. Then

$$p_i = \int_{a_i}^{b_i} \frac{1}{b_i - a_i} \int_t^{\infty} \mu e^{-\mu s}\, ds\, dt$$
$$= \int_{a_i}^{b_i} \frac{1}{b_i - a_i} e^{-\mu t}\, dt$$
$$= \frac{1}{b_i - a_i} \int_{a_i}^{b_i} e^{-\mu t}\, dt$$
$$= \frac{e^{-\mu a_i} - e^{-\mu b_i}}{\mu(b_i - a_i)}, \tag{22}$$

$$q_i = 1 - \frac{e^{-\mu a_i} - e^{-\mu b_i}}{\mu(b_i - a_i)}; \tag{23}$$

$$\frac{\partial p_i}{\partial \mu} = \frac{\partial q_i}{\partial \mu} = \frac{1}{b_i - a_i} \frac{-\mu(a_i e^{-\mu a_i} - b_i e^{-\mu b_i}) - (e^{-\mu a_i} - e^{-\mu b_i})}{\mu^2}$$

$$= -\left\{\frac{a_i e^{-\mu a_i} - b_i e^{-\mu b_i}}{\mu(b_i - a_i)} + \frac{p_i}{\mu}\right\}, \tag{24}$$

$$\frac{\partial^2 p_i}{\partial \mu^2} = -\frac{1}{(b_i - a_i)} \frac{\partial}{\partial \mu} \frac{a_i e^{-\mu a_i} - b_i e^{-\mu b_i}}{\mu} - \frac{\mu \frac{\partial p_i}{\partial \mu} - p_i}{\mu^2}$$

$$= -\frac{1}{b_i - a_i} \frac{-\mu(a_i^2 e^{-\mu a_i} - b_i^2 e^{-\mu b_i}) - (a_i e^{-\mu a_i} - b_i e^{-\mu b_i})}{\mu^2}$$

$$-\frac{1}{\mu}\frac{\partial p_i}{\partial \mu} + \frac{p_i}{\mu^2}$$

$$= \frac{a_i^2 e^{-\mu a_i} - b_i^2 e^{-\mu b_i}}{\mu(b_i - a_i)} + \frac{a_i e^{-\mu a_i} - b_i e^{-\mu b_i}}{\mu^2(b_i - a_i)} - \frac{1}{\mu}\frac{\partial p_i}{\partial \mu} + \frac{p_i}{\mu^2}$$

$$= \frac{a_i^2 e^{-\mu a_i} - b_i^2 e^{-\mu b_i}}{\mu(b_i - a_i)} - \frac{2}{\mu}\left\{\frac{\partial p_i}{\partial \mu}\right\}. \tag{25}$$

The asymptotic variance of the MLE of μ is obtained as

$$\text{AsVar}(\hat{\mu}) = \frac{1}{-E\left[\frac{\partial^2 \ell}{\partial \mu^2}\right]} \tag{26}$$

which is estimated by

$$\text{AsVar}(\hat{\mu}) = -\frac{1}{\partial^2 \ell / \partial \mu^2} \tag{27}$$

evaluated at $\mu = \hat{\mu}$.

The square root of the estimated asymptotic variance can be used as an estimate of the standard error of the estimate of μ. It provides a measure of the reliability of the estimate $\hat{\mu}$ based on the observed data. Since under mild regularity conditions the MLE has an asymptotically normal distribution, approximate 95 percent confidence bounds can be calculated by adding and substracting 1.96 $\sqrt{\text{AsVar}(\hat{\mu})}$ from the MLE $\hat{\mu}$.

The asymptotic variance of any function $g(\hat{\mu})$ can be obtained as

$$\text{AsVar}\,(g(\hat{\mu})) = \text{AsVar}\,(\mu)\{g'(\hat{\mu})\}^2. \tag{28}$$

It is well known [5, p. 43] that in the case of an exponential distribution with data that are not censored or grouped, $\hat{\mu}^{1/3}$ approaches normality more quickly (as the sampled size increases) than $\hat{\mu}$. Because of this, we will base estimates of confidence bounds of various quantities on the confidence bounds of $\hat{\mu}^{1/3}$ using the assumption of normality of $\hat{\mu}^{1/3}$. We obtain its asymptotic variance using (28) as

$$\text{AsVar}(\hat{\mu}^{1/3}) = \text{AsVar}(\hat{\mu})\frac{1}{9}\mu^{-4/3}. \tag{29}$$

In the next section, confidence bounds on $\hat{\mu}^{1/3}$ are transformed directly to obtain confidence bounds on related quantities. The reader interested in reviewing the properties of the maximum likelihood estimator should consult Cox and Hinkley [3, Ch. 9] or similar texts on statistics.

NUMERICAL RESULTS

Maximum likelihood estimates of the intensity function and the expected time to the next stage as well as upper and lower 95 percent confidence bounds were calculated using the methods described in the previous section. These calculations were carried out using both a midpoint departure assumption as well as a random censoring assumption. Table 2 indicates that the two assumptions produce virtually identical results. Consequently, we shall henceforth present results based on the midpoint method only.

TABLE 2

ESTIMATES AND 95% CONFIDENCE BOUNDS OF THE INTENSITY FUNCTION AND THE EXPECTED TIME IN STAGE

	Stage 1a (At-Risk)	Stage 1b (HIV+)	Stage 2a (LAS)	Stage 2b (ARC)	Stage 3 (AIDS)
A) Midpoint Method					
Intensity Function	0.45	0.86	0.53	0.30	1.1
Lower Confidence Bound	0.27	0.57	0.40	0.18	0.60
Upper Confidence Bound	0.69	1.2	0.70	0.46	1.8
Expected Time in Stage	2.2	1.2	1.9	3.4	0.93
Lower Confidence Bound	1.4	0.81	1.4	2.2	0.57
Upper Confidence Bound	3.7	1.7	2.5	5.7	1.7
B) Random Censoring Method					
Intensity Function	0.45	0.88	0.54	0.30	1.1
Lower Confidence Bound	0.27	0.58	0.40	0.18	0.61
Upper Confidence Bound	0.70	1.3	0.71	0.47	1.8
Expected Time in Stage	2.2	1.1	1.9	3.4	0.91
Lower Confidence Bound	1.4	0.79	1.4	2.1	0.55
Upper Confidence Bound	3.7	1.7	2.5	5.7	1.6

The expected time to the next stage is the average future length of time that any subject in a particular stage will wait before moving to the next stage. A consequence of the model is that this quantity does not depend upon the length of time a subject has been in the stage already; or stated equivalently, no aging of the subject occurs in any stage.

Table 3 gives the estimated cumulative distribution function and corresponding confidence limits for the time in any given stage. These numbers represent the proportion of persons who can be expected to have progressed to the next stage after 0.5, 1, 2, or 3 years in the stage. The estimates obtained by Cowell and Hoskins [2, Part 2, p. 25] are also given for comparative purposes. It should be noted that 11 out of 16 of their values fall within our confidence limits. Since we would expect about 15 out of 16 (that is, 95 percent) it would appear that our results are somewhat inconsistent with theirs. Furthermore, it should be noted that our estimates are all higher than those of Cowell and Hoskins.

TABLE 3

PROPORTION PROGRESSING TO NEXT STAGE IN SPECIFIED TIME

Stage	Time	Proportion	Lower Bound	Upper Bound	Cowell and Hoskins
1a	0.5	20%	13%	29%	—
(At-Risk)	1	36	24	50	—
	2	59	42	75	—
	3	74	55	87	—
1b	0.5	35	25	46	10%
(HIV+)	1	58	44	71	55
	2	82	68	92	75
	3	93	82	98	80
2a	0.5	23	18	30	15
(LAS)	1	44	33	50	41
	2	66	55	75	61
	3	80	70	88	75
2b	0.5	14	8	21	5
(ARC)	1	26	16	37	10
	2	45	30	60	51
	3	59	41	75	58
3	0.5	42	26	58	26*
(AIDS)	1	66	45	83	45*
	2	88	70	97	70*
	3	96	84	99	80*

*Cowell and Hoskins used data from the U.S. Center for Disease Control in lieu of the Frankfurt Study data to obtain the stage 3 values [2, Part 2, pp. 3 and 11].

Cowell and Hoskins [2, Part 2, p. 12] assumed the maximum length of observation periods throughout in order to offset the length of time from the

onset of the stage to entry into the study. This will have the effect of decreasing the intensity function and the proportion progressing to the next stage. This should explain the downward bias in their results from ours. It should be noted that our model is memoryless and requires no assumption about the time-in-stage before entering the study.

Table 4 gives the expected time to death of persons in any stage and 95 percent confidence bounds on these expected times. For computational reasons, the method used to calculate confidence bounds here is based on the assumption of asymptotic normality of these estimators rather than asymptotic normality of $\hat{\mu}^{1/3}$ as in the previous calculations. Consequently, these confidence bounds will be slightly inconsistent with those developed previously but should still give the reader some measure of the degree of reliability of the life expectancies.

TABLE 4

LIFE EXPECTANCY OF A PERSON IN ANY STAGE

Stage	Life Expectancy	Lower Bound	Upper Bound	Cowell and Hoskins*
1a (At-Risk)	9.6	7.5	12	—
1b (HIV+)	7.3	5.5	9.2	11.1
2a (LAS)	6.2	4.4	8.0	8.8
2b (ARC)	4.3	2.6	6.0	6.7
3 (AIDS)	0.93	0.44	1.4	2.1

*Obtained by addition of components found in Cowell and Hoskins [2, Part 2, p. 12].

As mentioned above, Cowell and Hoskins' methodology results in an upward bias in the life expectancies as well.

To this point we have not tested the validity of our constant intensity Markov process model. We do this by fitting or predicting the number of persons progressing to the next stage of the exposure base in Table 1A and comparing the results statistically with those of Table 1B. The results of these calculations are given in Table 5.

TABLE 5

PREDICTED (ACTUAL) PROGRESSIONS BASED ON THE MODEL

Observation Period	Stage 1a (At-Risk)	Stage 1b (HIV+)	Stage 2a (LAS)	Stage 2b (ARC)	Stage 3 (AIDS)
3-6 months	1.5(1)	2.5(1)	3.8(3)	0.8(0)	2.0(4)
6-12 months	4.0(6)	8.6(10)	16.9(20)	5.8(3)	5.0(6)
12-24 months	10.3(9)	14.5(15)	16.0(14)	7.1(10)	5.6(5)
24-36 months	2.0(2)	4.4(4)	14.0(14)	3.7(4)	0.9(0)

The fit appears to be remarkably good in the sense that it predicts the number of progressions accurately. A chi-squared statistic of 10.1 in comparison with a χ^2_{14} variable (14 degrees of freedom, since 20 cells and 5 parameters, $14 = 20 - 5 - 1$) at any reasonable significance level indicates that the model adequately describes the data.

This conclusion does not mean that the model used here precisely describes the physical phenomena underlying the progression of subjects through the various stages. The addition of further data in the near future could well require further refinements to the model. However, such refinements cannot be justified yet on the basis of the data available in this paper.

Table 6 presents a model multistage "life table" for persons who are in stage 1b (HIV+). At each duration, it gives the distribution by stage of a cohort of persons who were HIV+ initially. The differences between Cowell and Hoskins' methodology and ours are reflected in this table as well (see Cowell and Hoskins [2, Part 2, p. 26]).

TABLE 6

PERCENT DISTRIBUTION BY STAGE AND YEARS SINCE HIV INFECTION

Years since HIV Infection	Stage 1b (HIV+)	Stage 2a (LAS)	Stage 2b (ARC)	Stage 3 (AIDS)	Dead
0	100.0	0.0	0.0	0.0	0.0
0.5	64.9	30.5	4.4	0.2	0.0
1	42.2	43.1	13.2	1.2	0.4
1.5	27.4	45.9	22.6	2.8	1.4
2	17.8	43.4	30.7	4.7	3.4
2.5	11.5	38.7	36.7	6.7	6.5
3	7.5	33.1	40.5	8.3	10.5
3.5	4.9	27.6	42.5	9.6	15.4
4	3.2	22.6	42.8	10.5	20.8
4.5	2.1	18.3	42.0	11.0	26.7
5	1.3	14.6	40.3	11.1	32.6
6	0.6	9.1	35.3	10.6	44.4
7	0.2	5.6	29.6	9.4	55.2
8	0.1	3.4	24.0	8.0	64.6
9	0.0	2.0	19.1	6.5	72.4
10	0.0	1.2	14.9	5.2	78.7
11	0.0	0.7	11.5	4.1	83.7
12	0.0	0.4	8.8	3.2	87.6
13	0.0	0.2	6.7	2.5	90.6
14	0.0	0.1	5.1	1.9	92.9
15	0.0	0.1	3.8	1.4	94.7
16	0.0	0.1	2.8	1.1	96.0
17	0.0	0.0	2.2	0.8	97.0
18	0.0	0.0	1.6	0.6	97.8
19	0.0	0.0	1.2	0.5	98.3
20	0.0	0.0	0.9	0.3	98.8
21	0.0	0.0	0.7	0.3	99.1
22	0.0	0.0	0.5	0.2	99.3
23	0.0	0.0	0.4	0.1	99.5
24	0.0	0.0	0.3	0.1	99.6
25	0.0	0.0	0.2	0.1	99.7

OBSERVATIONS AND CONCLUSIONS

The purpose of this paper was twofold. First, the numerical results should be of interest to various parties, including insurers, dealing with the issue of AIDS. Second, the usefulness of simple parametric survival models was demonstrated. Their validity can be tested, standard errors and confidence bounds on parameters and related quantities can be calculated easily and they have intuitive appeal because of their inherent smoothness.

In the case discussed in this paper, five parameters were estimated using twenty independent pieces of data. In methods such as those used by Cowell and Hoskins, in effect, twenty quantities are estimated by twenty independent data points. This makes the results highly sensitive to the data. Parmetric models with an inherent smoothing function are more robust under small changes to the data. Furthermore, for this data set the parametric model implicitly projects beyond the longest observation period (three years).

The constant intensity model selected in this study is very simple. Although it is justified on the basis of the data provided, it is probably reasonable to expect that intensity functions would increase by duration-in-stage, that is, that subjects age or deteriorate making progression to the next stage more probable with duration in stage. This would require the introduction of more parameters. Furthermore, the issue of left-hand censoring becomes important and difficult to handle. How long a person has been in a particular stage prior to diagnosis has a direct effect on the likelihood function since it affects the level of the intensity function to be used at the time of diagnosis (entry into the study). Adding complexity to the model requires the use of more parameters in the model. As more data become available, it will be necessary to present the data in a format that will allow the user to extract information about the parameters.

Finally, a few comments regarding the format of the data in this study would be appropriate. By providing only information on the subjects' initial stage, no contribution is made to inferences about subsequent stages for persons who may have passed through several stages of the data. If this is done the standard errors of the estimates of the intensities of stages 1b, 2a, 2b and 3 can be discussed and the reliability of the various estimates increased. I hope that this will be done in subsequent reports.

REFERENCES

1. BRODT, H.R., HELM, E.B., WENER, A., JOETTEN, A., BERGMANN, L., KLUVER, A., AND STILLE W. "Spontanverlauf der LAV/HTLV-III-Infektion; Verlaufsbeobachtungen

bei Personen aus AIDS-Risikogruppen," *Deutsche Medizinische Wochenschrift, Stuttgart*, III (1986):1175–80.

2. COWELL, M.J. AND HOSKINS, W.H. "AIDS, HIV Mortality and Life Insurance." In *The Impact of AIDS on Life and Health Insurance Companies: A Guide for Practicing Actuaries, Report of the Society of Actuaries Task Force on AIDS*. Itasca, Ill.: Society of Actuaries, 1988.
3. COX, D.R. AND HINKLEY, D.V. *Theoretical Statistics*. London: Chapman and Hall, 1974.
4. EVERITT, B. AND HAND, D. *Finite Mixture Distributions*. London: Chapman and Hall, 1981.
5. KALBFLEISCH, J.G. *Probability and Statistical Inference*, Vol. 2, 2d ed. New York: Springer-Verlag, 1985.
6. REDFIELD, R., WRIGHT, D.C. AND TRAMONT, E.C. "The Walter Reed staging classification for HTLV-III/LAV Infection," *New England Journal of Medicine* 314, No. 2 (1986):131–32.

DISCUSSION OF PRECEDING PAPER

ELIAS S.W. SHIU:

Professor Panjer is to be congratulated for this timely and illuminating paper. I wish to follow up on the remark that "it is probably reasonable to expect that intensity functions would increase by duration-in-stage, that is, that subjects age or deteriorate making progression to the next stage more probable with duration in stage." As the mode of an exponential distribution is at zero, the model implies that, for a patient who has just entered stage j, it is *most probable* that he will enter stage $j+1$ *immediately*.

This observation about the exponential distribution assumption has been made elsewhere. In marketing analysis, it has been assumed that purchases of a particular brand of goods by a given consumer in successive time periods are independent and follow a Poisson distribution with a constant mean. Thus, interpurchase times are assumed to be exponential. Chatfield and Goodhardt [1, p. 828] criticize this assumption by pointing out that "it is in fact improbable that a buyer is most likely to buy again *immediately*." One may make a similar statement about claim occurrences in the classical collective risk model in which interclaim times are assumed to be exponentially distributed.

On December 10–11, 1987, the Council of Professional Associations on Federal Statistics convened a group of approximately 100 statisticians and data users in Bethesda, Md., for a symposium entitled "Federal Statistics on AIDS: Progress, Problems, Prognosis." Participants came from federal, state and local public health agencies, private research organizations, universities, and industry. At the end of the symposium, one observer noted: "Depending on one's perspective, the prognosis for quality statistics on AIDS is either optimistic or grim. On the bright side, a large number of first-rate epidemiologists, statisticians, and social scientists are hard at work trying to measure important dimensions of the AIDS epidemic and its consequences. On the other hand, we know so little about the transmission and development of the disease that the uncertainty associated with any forecasts is often as large as the forecasts themselves. This is rather frightening" [2, p. 8]. I would like to conclude this discussion by suggesting that, by the painstaking work of researchers such as Professor Panjer, we shall understand more and more about AIDS.

REFERENCES

1. CHATFIELD, C. AND GOODHART, G.J. "A Consumer Purchasing Model with Erlang Inter-Purchase Times," *Journal of the American Statistical Association* 68 (1973):828–35.
2. "Statistical Symposium on AIDS." *Chance: New Directions for Statistics and Computing*, 1 no. 2 (1988):7–8.

BEDA CHAN:

Professor Panjer is to be congratulated for an excellent piece of survival analysis illustrating the beauty of parsimonious parametric models. We would attempt to discuss the difference between the estimates of the current paper and those of Cowell and Hoskins.*

In Table 3, among estimates based on the Frankfurt Study, five out of twelve of Cowell and Hoskins' estimates fall outside the 95 percent confidence interval of the current paper; one would expect fewer than one out of twelve (that is, 5 percent). In all twelve, Cowell and Hoskins gave a lower proportion of progression to next stage and hence a higher expected time in one stage.

The following interesting property of the exponential distribution may help to explain the discrepancy. Let the service time of light bulbs be exponentially distributed with expected value 1 (year). If *all* light bulbs were dated when they were put on new and when they were dead and replaced, then the *average* service time is 1 (year), as expected. However, the expected service time of a randomly *inspected* light bulb is 2 (years): 1 for expected future lifetime because the exponential distribution is memoryless and because used is as good as new, and 1 for expected past lifetime by time reversal. The light bulbs that last longer are around longer and are more likely to be caught at inspection. This *average* lifetime versus *inspected* lifetime phenomenon could be the key.

Because the Cowell and Hoskins model considers duration in stage and because the Frankfurt Study inspected some but did not catch all HIV+ patients, their estimates may tend towards the *inspected* lifetime. In the current paper, the exponential model is assumed, μ is estimated, and $1/\mu$, as time in a stage, is *average* lifetime.

*COWELL, M.J. AND HOSKINS, W.H. "AIDS, HIV Mortality and Life Insurance." In *The Impact of AIDS on Life and Health Insurance Companies: A Guide for Practicing Actuaries, Report of the Society of Actuaries Task Force on AIDS*. Itasca, Ill.: Society of Actuaries, 1988.

J.C. MCKENZIE SMITH:

I congratulate the author for a very timely paper that also illustrates the use of Markov processes in an insurance setting. I wish to discuss an alternative to Equation (10) for obtaining Table 6.

We can define $p(t)$ as a vector of state occupancy probabilities [2] at time t and P as the constant matrix of transition probabilities as follows:

$$p(t) = \begin{Bmatrix} \text{probability of being in stage 1b at time } t \\ \text{probability of being in stage 2a at time } t \\ \text{probability of being in stage 2b at time } t \\ \text{probability of being in stage 3 at time } t \\ \text{probability of being dead at time } t \end{Bmatrix}$$

$$P = \begin{Bmatrix} -0.86 & 0.00 & 0.00 & 0.0 & 0.0 \\ 0.86 & -0.53 & 0.00 & 0.0 & 0.0 \\ 0.00 & 0.53 & -0.30 & 0.0 & 0.0 \\ 0.00 & 0.00 & 0.30 & -1.1 & 0.0 \\ 0.00 & 0.00 & 0.00 & 1.1 & 0.0 \end{Bmatrix}$$

The diagonal elements of P are the negatives of the intensities shown in Table 2(A) to two significant digits of accuracy. The only nonzero nondiagonal elements are those corresponding to transition to the next stage of the disease; this is consistent with the author's treatment and the underlying data. The columns sum to zero, indicating that no one is lost from the system. The elements in the last column are all zero because death is an "absorbing" state. Finally, the elements above the diagonal are all zero because it was assumed that no recoveries occur.

The vector $p(t)$ obeys a differential equation:

$$dp(t)/dt = P\,p(t) \tag{A}$$

The solution to (A), as discussed in [1,2], is

$$p(t) = F(t,0)\,p(0) \tag{B}$$

where

$$\begin{aligned} F(t,s) &= I + (t-s)P + [(t-s)P]^2/2! + \ldots + [(t-s)P]^n/n! \\ &\quad + \ldots, \\ (t-s)P &\quad \text{means multiplication of matrix } P \text{ by scalar } t-s, \\ I &= \text{the identity matrix,} \end{aligned} \tag{C}$$

and the powers are in the context of matrix multiplication. It turns out that $F(t,s)$ has the following properties:

(1) $F(t,t) = I$ for all $t \geqq 0$.
(2) $F(s,t)$ = matrix inverse of $F(t,s)$ for $0 \leqq s \leqq t$.
(3) $F(t,s) = F^{2(t-s)}$ where $F = F(0.5,0)$.
(4) $F(t,s)\,F(s,r) = F(t,r)$ for $r,s,t \geqq 0$ and therefore
(5) $p(t) = F(t,s)\,p(s)$ for all $s \geqq 0 \geqq t$.

From these results, it follows that we can calculate $F = F(0.5,0)$ by using as many terms as necessary for convergence in (C):

$$F = \begin{Bmatrix} 0.6493 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.3047 & 0.7654 & 0.0000 & 0.0000 & 0.0000 \\ 0.0436 & 0.2174 & 0.8624 & 0.0000 & 0.0000 \\ 0.0020 & 0.0144 & 0.1057 & 0.5837 & 0.0000 \\ 0.0003 & 0.0028 & 0.0319 & 0.4163 & 1.0000 \end{Bmatrix}$$

The state occupancy probabilities can be calculated at six-month intervals recursively as follows:

$$p(0) = \begin{Bmatrix} 100.0\% \\ 0.0\% \\ 0.0\% \\ 0.0\% \\ 0.0\% \end{Bmatrix}; \quad p(0.5) = F\,p(0) = \begin{Bmatrix} 64.9\% \\ 30.5\% \\ 4.4\% \\ 0.2\% \\ 0.0\% \end{Bmatrix};$$

$$p(1.0) = F\,p(0.5) = \begin{Bmatrix} 42.2\% \\ 43.1\% \\ 13.2\% \\ 1.2\% \\ 0.4\% \end{Bmatrix}; \text{ etc.}$$

In general, $p(t) = F\,p(t-0.5)$. These numbers are the same as in Table 6 except that these are probabilities that sum to 100 percent, whereas the Table 6 numbers are expectations that sum to 100.

The only nonroutine part of this approach is to determine the form of P. The rest is straightforward number-crunching. These techniques are described in [1,2] in more detail. Because this is a recursive method, it is important to carry as many decimal places as possible to avoid the accumulation of rounding errors. The results in this discussion were shown to one decimal place to be consistent with the author's presentation.

In general, the state occupancy probabilities at n half-years for a life in state i at time 0 are displayed in the ith column vector of the nth matrix power of F, which may be calculated by taking successive powers of $F = F(0.5,0)$.

Multistate reserves could be calculated by using the results of this paper and the techniques discussed in [3]. Such reserves could even incorporate recovery assumptions if these became appropriate. However, an insurance company would not have sufficiently good information on the number of its covered lives in each state to make use of reserves that assume a knowledge of the stage of the disease. Reserves based on the knowledge that a person has been infected combined with an estimate of face amounts on infected lives may be a practical alternative.

One final caveat is not to try to apply (C) to cases in which P varies with time. This case requires special handling, which is discussed in [1].

REFERENCES

1. BROCKETT, R.W. *Finite Dimensional Linear Systems*. New York, London, Sydney, Toronto: John Wiley and Sons, Inc., 1970.
2. KLEINROCK, L. *Queueing Systems*. Vol. I, *Theory*. New York, London, Sydney, Toronto: John Wiley and Sons, 1975.
3. SMITH, J.C. McK. "A Ballistic Approach to Actuarial Problems," *TSA* XXXVI (1984): 501–22.

MARK D.J. EVANS:

Professor Panjer has developed a statistical model for AIDS survival analysis that addresses some problems with the Cowell and Hoskins approach, increasing the usefulness of their extensive research in this area. There are a few areas of Professor Panjer's analysis that merit further reflection.

The most important concern is the data used for stage 3. Panjer uses the Frankfurt data, which studied only 23 individuals. Cowell and Hoskins use the CDC data, which not only contain more individuals (33,482) but also tabulate the data by date of diagnosis rather than by time in stage. Use of the CDC data is clearly preferable.

The CDC stage 3 progression information is quite reasonably modeled by Cowell and Hoskins; it does not fit the exponential distribution used by Panjer. Unfortunately, this makes inapplicable some of the convenient relationships developed by Panjer. This could be addressed by assuming (somewhat inaccurately) that the CDC data fit an exponential distribution. The CDC data imply a stage 3 life expectancy of 2.1 years, which implies an intensity function of 0.48, as opposed to the Frankfurt-based value of

1.1. Use of the stage 3 data from the CDC study would increase the life expectancy estimate calculated by Panjer by about 1.2 years.

There is a peculiarity in the CDC data, which, although it does not significantly impact any results, does pose some questions. The Cowell and Hoskins stage 3 model predicts cumulative mortality of 95 percent and 96 percent at 8 years and 8.5 years, respectively. The actual results are as follows:

Observation Period	Patients Observed	Patients Expired	Percentage
8 Years	4	2	50%
8.5 Years	10	6	60%

At first, this might appear to be a statistical fluke caused by a small amount of data. However, constructing binomial distributions based upon the predicted percentage of 95 percent and 96 percent demonstrates that if these are the true underlying probabilities, the chances of getting no more than two deaths at year 8 and no more than six deaths at year 8.5 are about 1 in 100,000,000. Perhaps there is some problem with the data at these extended periods or the earlier victims of AIDS were infected with a less devastating strain.

The stage 1b data from the Frankfurt Study also contain some interesting characteristics. If we use least squares to fit the data to the function $B_0 + B_1 \times 1/t$, where t represents time in stage exposed to the interval midpoint, we get a "reciprocal" relationship. This results in the following:

$B_0 = 0.946$
$B_1 = -0.309$

PERCENTAGE WHOSE HEALTH WORSENED

t	Observed	Exponential Fit	Reciprocal Fit
3/8	0.11	0.28	0.12
3/4	0.56	0.48	0.53
3/2	0.75	0.72	0.74
5/2	0.80	0.88	0.82

F test:

Exponential versus reciprocal fit produces F statistic of 23.2, which is greater than $F_{0.05;3,2}(=19.2)$.

Note that technically this model should be exposed as follows:

$$F(t) = 0,\ 0 \le t \le 0.327$$

$$F(t) = 0.946 - \frac{0.309}{t},\quad 0.327 < t$$

These results suggest three conclusions. First, the exponential distribution is not the best model for these data. Second, the data are not significantly compromised by left-hand considerations. Third, this model suggests that 5 percent of those in stage 1b will never progress. (The 95 percent confidence level for 0.946 excludes the value 1, but 1 is included at the 97.5 percent level.)

The data for stages 2a and 2b do not appear to suggest that anything other than Professor Panjer's approach would be appropriate.

In conclusion, although Professor Panjer has made some important observations about the limitations of the techniques used by Cowell and Hoskins and presented a model that resolves these problems, his model clearly needs to be modified to incorporate the superior CDC stage 3 data. Also, this model should make better use of the Frankfurt stage 1b data. Unfortunately, these findings introduce complications into Panjer's development of the distribution resulting from the combination of the various stages.

To overcome the difficulties, I developed a computer model by using Markov chain techniques, similar to the approach used by Cowell and Hoskins. The model generates monthly progressions and makes approximate adjustments to compensate for the half-month average implicit progression delay.

The table on page 538 is an attempt to reproduce Panjer's Table 6 by using the computer model. Note that this computer model produces results similar to Panjer's. It produces an HIV+ life expectancy that is within 0.1 years of Panjer's.

Next I reran the model, replacing the assumptions for stage 1b and stage 3 as documented below:

Models for Each Stage

Stage 1b:

$$F(t) = 0,\qquad 0 \le t \le 0.327$$

PERCENTAGE DISTRIBUTION BY STAGE AND YEARS SINCE HIV INFECTION

Years Since HIV Infection	Stage 1b (HIV+)	Stage 2a (LAS)	Stage 2b (ARC)	Stage 3 (AIDS)	Dead
0	100.0	0.0	0.0	0.0	0.0
0.5	65.1	32.0	4.9	0.2	0.0
1	42.3	43.6	13.8	1.3	0.4
1.5	27.5	45.9	23.1	2.9	1.5
2	17.9	43.3	30.9	4.9	3.7
2.5	11.6	38.5	36.7	6.8	6.9
3	7.6	32.9	40.3	8.4	11.0
3.5	4.9	27.5	42.1	9.6	16.0
4	3.2	22.5	42.4	10.4	21.5
4.5	2.1	18.2	41.5	10.9	27.4
5	1.4	14.6	39.7	11.0	33.4
6	0.6	9.2	34.7	10.4	45.2
7	0.2	5.6	29.0	9.2	56.0
8	0.1	3.5	23.5	7.7	65.3
9	0.0	2.1	18.6	6.3	73.0
10	0.0	1.2	14.5	5.0	79.2
11	0.0	0.7	11.2	3.9	84.1
12	0.0	0.4	8.5	3.0	88.0
13	0.0	0.2	6.5	2.3	90.9
14	0.0	0.1	4.9	1.8	93.2
15	0.0	0.0	3.7	1.3	94.9
16	0.0	0.0	2.8	1.0	96.2
17	0.0	0.0	2.1	0.8	97.2
18	0.0	0.0	1.5	0.6	97.9
19	0.0	0.0	1.1	0.4	98.4
20	0.0	0.0	0.8	0.3	98.8
21	0.0	0.0	0.6	0.2	99.1
22	0.0	0.0	0.5	0.2	99.4
23	0.0	0.0	0.3	0.1	99.5
24	0.0	0.0	0.2	0.1	99.7
25	0.0	0.0	0.2	0.1	99.8

$$F(t) = 0.946 \frac{0.309}{t}, \quad 0.327 < t < 21$$

$$F(t) = 1 \qquad 21 \leq t^*$$

Stage 2a:

$$F(t) = 1 - e^{-0.53t}$$

Stage 2b:

$$F(t) = 1 - e^{-0.30t}$$

*Normal male age 35 mortality would reduce the population by 6.8 percent in 21 years, so this point was chosen for cutoff. This assumes probabilities are additive, although one could argue they are probably multiplicative.

Stage 3:

$$F(t) = 1 - 0.55^t \qquad 0 \le t \le 2$$

$$F(t) = 1 - 0.55^2 \times 0.65^{(t-2)} \qquad 2 \le t \le 3$$

$$F(t) = 1 - 0.55^2 \times 0.65 \times 0.75^{(t-3)} \; 3 \le t$$

The following results were produced by this model:

PERCENTAGE DISTRIBUTION BY STAGE AND YEARS SINCE HIV INFECTION

Years Since HIV Infection	Stage 1b (HIV+)	Stage 2a (LAS)	Stage 2b (ARC)	Stage 3 (AIDS)	Dead
0	100.0	0.0	0.0	0.0	0.0
0.5	67.2	34.9	2.3	0.1	0.0
1	36.3	50.3	13.3	1.1	0.1
1.5	26.0	46.9	23.6	3.2	0.8
2	20.9	40.3	31.0	6.0	2.1
2.5	17.8	33.5	35.8	8.8	4.3
3	15.7	27.4	38.2	11.4	7.4
3.5	14.2	22.3	39.0	13.5	11.0
4	13.1	18.1	38.5	15.2	15.2
4.5	12.3	14.6	37.1	16.4	19.7
5	11.6	11.8	35.2	17.1	24.4
6	10.6	7.7	30.4	17.7	33.7
7	9.8	5.1	25.4	17.1	42.6
8	9.3	3.4	20.7	16.0	50.6
9	8.8	2.3	16.6	14.5	57.8
10	8.5	1.6	13.2	12.8	63.9
11	8.2	1.2	10.4	11.1	69.1
12	8.0	0.9	8.2	9.5	73.5
13	7.8	0.7	6.4	8.1	77.1
14	7.6	0.5	5.1	6.8	80.1
15	7.5	0.4	4.0	5.7	82.6
16	7.3	0.3	3.1	4.7	84.6
17	7.2	0.3	2.4	3.9	86.2
18	7.1	0.2	1.9	3.2	87.6
19	7.0	0.2	1.5	2.6	88.6
20	6.9	0.2	1.2	2.1	89.5
21	0.0	6.7	1.3	1.7	90.2
22	0.0	4.0	3.3	1.8	90.9
23	0.0	2.3	3.8	2.0	91.8
24	0.0	1.4	3.7	2.2	92.8
25	0.0	0.8	3.2	2.1	93.9

As mentioned earlier, use of the CDC stage 3 data increases life expectancy by about 1.2 years, while use of the reciprocal model results in an additional 1.6 years increase, giving an HIV+ life expectancy of 10 years.

(AUTHOR'S REVIEW OF DISCUSSION)

HARRY H. PANJER:

I wish to thank each of the discussants for their individual contributions. Professor Shiu points out the monotonicity of the probability density function of the exponential distribution. Unfortunately, the choice of a more complex model cannot be justified on the basis of the data, because the exponential models cannot be rejected either simultaneously for all stages combined (as is discussed following Table 5) or independently for each stage, as can be verified by the reader from Table 5 for each stage separately by using a χ^2 statistic with two degrees of freedom. The exponential distribution should be considered a starting point only. As more data emerge, models with more parameters can be justified.

Professor Chan points out the well-known "inspection paradox," which may be alternatively described as follows. Suppose that buses arrive at a bus stop on average 10 minutes apart (in accordance with a homogeneous Poisson process). If an individual arrives at the bus stop, the expected time to the next bus is *not* five minutes. It is still 10 minutes. The explanation is simple. The individual is more likely to arrive at the bus stop during a long interarrival time than a short one, because the long interarrival times take up more of the time scale than the short ones. In the context of our analysis of the data in the paper, persons who have short times in stage j are more likely to enter the study after stage j has passed. They may contribute to a subsequent stage in which the same arguments apply. Consequently, the individuals involved in the study will have a longer expected time in the initial stage (prior and subsequent to entry) than persons not in the study or persons entering in an earlier stage.

Mr. Smith provides an elegant alternative computational tool for evaluating various probabilities. These may be of special interest to readers.

Mr. Evans correctly points out that the CDC data for stage 3 are much better than the small data set in the Frankfurt Study. The only purpose of our study was to analyze the Frankfurt data. The Frankfurt data are useful primarily for information about the incubation period running from infection (stage 1b) to diagnosis of AIDS (stage 3).

He also suggests a two-parameter model for matching the distribution function. His model apparently fits better, although he does not recognize the sample sizes in the different exposure periods in his method of fitting. It should be pointed out that *any* model that includes the exponential model as a special case will also provide a better fit than the exponential. However,

the principle of parsimony dictates that the data set is not sufficiently large to justify a model that is more complicated than the exponential.

I would like to express my thanks to the reviewers and the many others who have expressed interest in this paper. I am finally most indebted to Mike Cowell, who prodded me (however gently) to do this study and to carry on further studies associated with this tragic epidemic of AIDS.