J. Hyg., Camb. (1979), **83**, 77
Printed in Great Britain

Estimating the duration of latency and survival time of snails with schistosomiasis

By M. J. GODDARD*

London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT

(*Received* 29 *August* 1978)

INTRODUCTION

The use of quantitative methods in all aspects of epidemiology is now commonplace. The study of schistosomiasis has certainly benefited from this trend. The spectrum of uses of quantitative techniques includes nearly all aspects of the study of this disease. Indeed, the use of these measures has produced a desire to describe the transmission of the disease by means of purely mathematical relationships. Even this area is of considerable size and diversity as any reader of a recent article by J. E. Cohen (1977) will find. As these tools develop, there need to be developed statistical methods to evaluate numerical parameters required by a mathematical description. Statistical methodology to evaluate parameters need not benefit the modellers only, as results will be of interest to epidemiologists and biologists as well.

The aim of this article is to describe a statistical technique applicable to the estimation of the survival of snails transmitting schistosomiasis. The basic model is a modification of a technique developed for use in studying survival times of cancer patients. While the methodology is only slightly altered from the cancer application, the prime purpose herein is to suggest its relevance to helminthologists, and to describe the way it is used and methods of testing the assumptions that are needed. The techniques are taken from different sections of the statistical literature, and one purpose is to draw these together.

The snail is a necessary intermediate vector in the transmission of schistosomiasis. Infective agents in the local water (called miracidia) can penetrate snails, and only a small proportion of the miracidia develop into mother sporocysts. These contribute to the transmission of infection by releasing cercariae, another infective aquatic form, which are capable of infecting human hosts. Snails which are infected but not releasing cercaria are termed 'prepatent', and are said to be in a 'latent stage' or 'latency'. Once cercarial release begins it continues until the death of the snail, during which time it is described as 'patent' or 'patently infected'. A thorough description of this and other aspects of transmission are found in the book by Jordan & Webbe (1969).

The model described below monitors the time to death of prepatent and patent snails, as well as the duration of latency following penetration. Estimates of the

* Present address: Unité de Recherches Statistiques, Institut Gustave-Roussy, 16 bis Avenue Paul-Vaillant-Couturier, 94800 Villejuif, France.

0022-1724/79/8 0103-1978 $01.00

parameter relating to the death rate of snails have been obtained by Sturrock & Webbe (1971). Their approach followed a catalytic model, as developed in a monograph by Muench (1959). On the assumption that the population dynamics of a colony of snails were stable and that age could be reliably estimated from the size of snails, catalytic curves were fitted by the method of moments and estimates of the death-rate obtained. This approach was modified by J. E. Cohen (1973), who made allowance for a possibly different rate for infected snails. This altered model was subsequently fitted by Sturrock, Cohen & Webbe (1975).

Leaving topics related to schistosomiasis aside momentarily, let us turn our attention to the measurement of survival in general. Study of this type of data probably grew first from industrial situations in which interest centred on the reliability of components and the time to failure of equipment was studied. Survival time data now are frequently encountered in clinical trials, particularly those measuring the effect of chemotherapy in cancer studies. The method described below modifies a model developed for cancer studies and, in order to study the validity of assumptions, draws on graphical techniques developed for industrial data.

In the next section the design of an experiment to determine the parameters of interest, with the keeping of records, is described. Section 3 contains the basic model and the estimation procedure. A hypothetical example, introduced in Section 2, is analysed. Assumptions made in formulating the model in Section 3 are tested by methods in Section 4. In the fifth section two embellishments on the original model are described with the appropriate estimation procedures and Section 6 is a summary.

2. EXPERIMENTAL PROCEDURE

The technique differs from the method of catalytic models in being an *in vitro* approach. A series of test tubes or microhabitats are arranged and snails are singly introduced to each. In the simplest case one would seek to have snails of one age and that being nearest the time they are first susceptible to miracidial penetration. (Below, an allowance can be made in the model for differing ages.)

For this model specifically, all snails should be exposed to sufficient miracidia to ensure each is infected. While we are primarily interested in the behaviour of infected snails, it is useful to have a parallel experiment on uninfected snails as a 'control'. In these circumstances randomization of all snails into the two groups is recommended. In the remainder of this report no mention will be made of estimation of parameters for the control population. The reader is referred to chapter 3 of the book by Gross & Clark (1975) for details.

Snails are uniquely numbered and on each day note is taken whether or not death has occurred and whether or not the latent period is over. Snails lost for unnatural reasons, such as dropped test tubes or unrelated infections, should be noted on the day lost. Note that allowance is made for snails not yet expired at the termination of the experiment. In Table 1 are presented some hypothetical data which will serve as an example for the technique of the next section.

Table 1. *Results of a hypothetical experiment*

Snail number	Day of death or lost	Day of patency	Died (= 1) or not dead (= 0)	Model variables			
				a	b	U	V
1	45	37	1	1	1	45	37
2	41	36	1	1	1	41	36
3	67	35	1	1	1	67	35
4	29	—	1	1	0	29	29
5	3	—	0*	0	0	3	3
6	19	—	0†	0	0	19	19
7	1	—	1	1	0	1	1
8	22	—	1	1	0	22	22
9	5	—	1	1	0	5	5
10	61	31	1	1	1	61	31
11	39	34	1	1	1	39	34
12	49	36	1	1	1	49	36
13	11	—	1	1	0	11	11
14	27	—	1	1	0	27	27
15	54	36	1	1	1	54	36
16	33	31	1	1	1	33	31
17	35	—	1	1	0	35	35
18	36	36	1	1	1	36	36
19	55	43	1	1	1	55	43
20	13	—	0†	0	0	13	13
21	17	—	1	1	0	17	17
22	57	39	1	1	1	57	39
23	26	—	0†	0	0	26	26
24	74	34	1	1	1	74	34
25	27	—	1	1	0	27	27
26	36	—	1	1	0	36	36
27	35	30	1	1	1	35	30
28	33	—	1	1	0	33	33
29	8	—	0*	0	0	8	8
30	15	—	1	1	0	15	15
31	42	35	1	1	1	42	35
32	5	—	1	1	0	5	5
33	10	—	1	1	0	10	10
34	12	—	1	1	0	12	12
35	16	—	0*	0	0	16	16

* Test tubes dropped.
† Snails infected from other organism.

3. PRIMARY MODEL

This model is based largely on one described by Lagakos (1976) in a cancer setting. Of interest were the time to death and the time to progression of the tumour. The current model studies the time to death of snails and the time until patency occurs.

For any snail, one of four possible events will be observed. These are:

E_1 Experiment terminated before snail death, or loss of snail for unnatural reason at time w_i, before patency.

E_2 Experiment terminated before snail death, or loss of snail for unnatural reason at time w_i, after patency at time p_i.

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

E_3 Snail death at t_i, after patency at time p_i.

E_4 Snail death at t_i, before patency.

Three random variables, X_1, X_2 and X_3, are introduced by Lagakos to model these events. From the start a snail is exposed to the competing risks of death (in a latent stage) or commencement of patency. The first random variable (or process), X_1, describes the time to death in latent snails and the second process, X_2, the time to patency. The third variable, X_3, models the time to death following patency. In the Lagakos formulation all three random variables were exponentially distributed with parameters λ_1, λ_2 and λ_3 respectively.

For present purposes it is convenient to retain the assumption that the time to death of snails is exponentially distributed. Thus X_1 and X_3 are assumed to be exponentially distributed with parameters λ_1 and λ_3 respectively. However, it does not seem justified to assume the duration of latency (X_2) is exponentially distributed. An alternative worthy of consideration might be the three parameter gamma density (see, for example, Harter & Moore, 1965). For present purposes a normal approximation is chosen to avoid considerable calculation. This distribution is a special case of the three-parameter gamma as one parameter (frequently referred to as the 'order') tends to infinity. Unfortunately, whereas the three parameter gamma density applies to $X_2 \geqslant 0$, by using the normal approximation, for some values of μ and σ^2, values of $X_2 < 0$ have a large probability. It is thought that in this application $\mu \gg \sigma^2$ and the tail of the distribution for which $\Pr(X_2 < 0)$ is negligible. Thus for the second process (of duration of latency) the random variable X_2 is normally distributed with mean μ and variance σ^2. (Wishing to keep present notation as similar as feasible to that of Lagakos, there will be no references to λ_2 and λ_3 will retain its earlier meaning. In Section 4 the four symbols θ_1, θ_3, γ_1 and γ_3 are introduced with θ_2 and γ_2 omitted for the same reason.)

This is just one possible generalization of the Lagakos model. Much more extensive generalizations are being studied by S. M. Gore in Aberdeen. These involve the use of Weibull densities for all three processes and, in particular, the introduction of time-dependence between stages. These models are for use in more complicated clinical trial settings while the simplified extensions contained herein are primarily aimed at studying snail life-span and duration of patency.

Observations on the snails are summarized by a 4-tuple (U_i, a_i, V_i, b_i) as follows. The variables a_i and b_i are indicators. Snails not dead at the time of analysis or lost for unnatural reasons yield 'censored' observations and for these snails $a_i = 0$. Snails that die in the experiment have associated $a_i = 1$. Snails for whom the time of patency was observed have $b_i = 1$, and for those with patency not observed (because time of patency is greater than time to death or loss) are censored and $b_i = 0$. The time of death or censoring is represented by U_i, and that of patency, or death (or loss) if earlier is represented by V_i. For example, Table 1 also displays the appropriate values of (U_i, a_i, V_i, b_i).

Estimation of the parameters λ_1 and λ_3 follows Lagakos' method and is reviewed here for completeness. Estimation of the parameters μ and σ^2 involves the solution

of maximum-likelihood equations for censored normal observations. The equations have been obtained by A. C. Cohen (1963) and notation here will resemble his as closely as possible.

Let

$$\xi_i = (V_i - \mu)/\sigma,$$

$$\phi(t) = (\sqrt{2\pi})^{-1} \exp(-t^2/2)$$

and

$$F(\xi_i) = \int_{-\infty}^{\xi_i} \phi(t)dt.$$

Ignoring constants, the logarithm of probability densities of the four events are:

$$\begin{aligned} \log P(E_1) &= -\lambda_1 U_i + \log\{1 - F(\xi_i)\} \\ &= -\lambda_1 V_i + \log\{1 - F(\xi_i)\} \quad \text{as} \quad U_i = V_i \quad \text{if } E_1 \text{ occurs} \end{aligned} \tag{3.1}$$

$$\log P(E_2) = -\lambda_1 V_i - \lambda_3(U_i - V_i) - \log\sigma - \tfrac{1}{2}\xi_i^2 \tag{3.2}$$

$$\log P(E_3) = -\lambda_1 V_i - \lambda_3(U_i - V_i) + \log\lambda_3 - \log\sigma - \tfrac{1}{2}\xi_i^2 \tag{3.3}$$

$$\begin{aligned} \log P(E_4) &= -\lambda_1 U_i + \log\lambda_1 + \log\{1 - F(\xi_i)\} \\ &= -\lambda_1 V_i + \log\lambda_1 + \log\{1 - F(\xi_i)\} \quad \text{as} \quad U_i = V_i \quad \text{if } E_4 \text{ occurs.} \end{aligned} \tag{3.4}$$

Writing $U_\cdot = \Sigma U_i$ and $V_\cdot = \Sigma V_i$, and n as the number of observations (snails), and using the indicators a_i and b_i, then the log likelihood ignoring constants is:

$$\begin{aligned} L = &-\lambda_1 V_\cdot - \lambda_3(U_\cdot - V_\cdot) + \log\lambda_1 \Sigma a_i(1 - b_i) + \log\lambda_3 \Sigma a_i b_i \\ &- \Sigma[(b_i/2)(\log\sigma^2 + \xi_i^2) - (1 - b_i)\log\{1 - F(\xi_i)\}]. \end{aligned} \tag{3.5}$$

(All summations without indices throughout this article are over the sample, that is Σ stands for $\sum_{i=1}^{n}$.)

Maximum-likelihood equations are obtained as usual using

$$\frac{\partial L}{\partial \lambda_1} = \frac{\Sigma a_i(1 - b_i)}{\lambda_1} - V_\cdot, \tag{3.6}$$

$$\frac{\partial L}{\partial \lambda_3} = \frac{\Sigma a_i b_i}{\lambda_3} - (U_\cdot - V_\cdot), \tag{3.7}$$

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma}[\Sigma\{b_i\xi_i + (1 - b_i)Z_i\}], \tag{3.8}$$

$$\frac{\partial L}{\partial \sigma} = \frac{1}{\sigma}[\Sigma\{b_i(\xi_i^2 - 1) + (1 - b_i)\xi_i Z_i\}], \tag{3.9}$$

where Z_i is the inverse of Mill's ratio or the 'hazard' function for the standard normal distribution:

$$Z_i = \frac{\phi(\xi_i)}{1 - F(\xi_i)}.$$

Estimates of λ_1 and λ_3 are directly obtained, as described by Lagakos:

$$\hat{\lambda}_1 = \frac{\Sigma a_i(1 - b_i)}{V_\cdot}, \quad \hat{\lambda}_3 = \frac{\Sigma a_i b_i}{(U_\cdot - V_\cdot)}.$$

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

Estimates of μ and σ are obtained iteratively by means of the Newton-Raphson technique as described in Cohen (1963). Again, for completeness, the derivatives are:

$$\frac{\partial^2 L}{\partial \lambda_1^2} = -\frac{\Sigma a_i(1-b_i)}{\lambda_1^2}, \tag{3.10}$$

$$\frac{\partial^2 L}{\partial \lambda_3^2} = -\frac{\Sigma a_i b_i}{\lambda_3^2} \tag{3.11}$$

$$\frac{\lambda^2 L}{\partial \lambda_1 \partial \lambda_3} = \frac{\partial^2 L}{\partial \lambda_1 \partial \mu} = \frac{\partial^2 L}{\partial \lambda_1 \partial \sigma} = \frac{\partial^2 L}{\partial \lambda_3 \partial \mu} = \frac{\partial^2 L}{\partial \lambda_3 \partial \sigma} = 0$$

$$\frac{\partial^2 L}{\partial \mu^2} = -\frac{1}{\sigma^2}\Sigma\{b_i + (1-b_i)A_i\}, \tag{3.12}$$

$$\frac{\partial^2 L}{\partial \mu \partial \sigma} = \frac{1}{\sigma^2}\Sigma\{2b_i\xi_i + (1-b_i)B_i\}, \tag{3.13}$$

$$\frac{\partial^2 L}{\partial \sigma^2} = -\frac{1}{\sigma^2}\Sigma[b_i(3\xi_i^2 - 1) + (1-b_i)C_i\}, \tag{3.14}$$

with

$$A_i = Z_i(Z_i - \xi_i), \tag{3.15}$$

$$B_i = Z_i + \xi_i A_i, \tag{3.16}$$

$$C_i = \xi_i(Z_i + B_i). \tag{3.17}$$

Thus, the approximate variance of $\hat{\lambda}_1$ is $\lambda_1^2/[\Sigma a_i(1-b_i)]$ and the approximate variance of $\hat{\lambda}_3$ is $\lambda_3^2/\Sigma a_i b_i$. The variance covariance matrix of $(\hat{\mu},\hat{\sigma})$ is obtained iteratively, the inverse of

$$-\begin{pmatrix} \dfrac{\partial^2 L}{\partial \mu^2} & \dfrac{\partial^2 L}{\partial \mu \partial \sigma} \\ \dfrac{\partial_2 L}{\partial \mu \partial \sigma} & \dfrac{\partial^2 L}{\partial \sigma^2} \end{pmatrix}$$

at $(\hat{\mu},\hat{\sigma})$.

For the example in Section 2, we have $U_\cdot = 1058$, $V_\cdot = 863$ with $\Sigma a_i(1-b_i) = 15$ and $\Sigma a_i b_i = 14$. Thus

$$\hat{\lambda}_1 = 15/863 = 0{\cdot}01738 \quad \text{and} \quad V(\hat{\lambda}_1) \simeq \Sigma a_i(1-b_i)/V_\cdot^2 = 2{\cdot}01405 \times 10^{-5}$$

and

$$\hat{\lambda}_3 = 14/(1058\text{–}863) = 0{\cdot}07179 \quad \text{and} \quad V(\hat{\lambda}_3) \simeq \Sigma a_i b_i/(U_\cdot - V_\cdot)^2 = 3{\cdot}68179 \times 10^{-4}.$$

Iterative solution for $\hat{\mu}$ and $\hat{\sigma}$ gives

$$\hat{\mu} = 35{\cdot}66 \quad \text{with} \quad V(\hat{\mu}) \simeq 0{\cdot}65993,$$

$$\hat{\sigma} = 3{\cdot}2066 \quad \text{with} \quad V(\hat{\sigma}) \simeq 0{\cdot}35078,$$

and cov $(\hat{\mu},\hat{\sigma}) \simeq -0{\cdot}06123$. The mean life-span of a snail in the absence of the competing risk of patency is $1/0{\cdot}01738 = 57{\cdot}5$ days. For a snail just infected, the mean duration of life is the expectation of the minimum of X_1 and $X_2 + X_3$ which is less than $57{\cdot}5$ or $35{\cdot}66 + (1/0{\cdot}07179) = 49{\cdot}6$ days. Once a snail becomes patent

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

the mean duration of life, $E(X_3)$, is 13·9 days. The average duration of latency is 35·7 days with variance $(3{\cdot}2066)^2 = 10{\cdot}28$ days^2. (Again, these data are hypothetical and no biological conclusions should be drawn from these results. They were generated with underlying values $\lambda_1 = 0{\cdot}01$, $\lambda_3 = 0{\cdot}05$, $\mu = 35$ and $\sigma^2 = 5$ or $\sigma = 2{\cdot}236$. Censoring was independent and exponential with $\lambda = 0{\cdot}025$.) A FORTRAN subroutine has been written and is available to anyone wishing to use this method on experimental data.

4. TESTING OF DISTRIBUTIONAL ASSUMPTIONS

The model of Section 3 has required the assumptions that (1) the time to death of prepatent snails is exponential, (2) the duration of latency is normally distributed and (3) the time to death after patency is exponential. The exponential assumptions simplify both the algebraic treatment and interpretation. It does seem appropriate to investigate the validity of the assumptions made. The methods described herein are drawn entirely from a paper by Nelson (1972) who detailed graphical tools in an industrial setting.

Tests of the three assumptions mentioned above are performed visually from plots of the time of event (such as death or patency) and the cumulative hazard function. The hazard function (or age-specific failure rate) at time t gives the instantaneous death-rate of a snail known to be of age t. The cumulative hazard does not have as comprehensible a meaning, but provides a better visual test. The calculation of the hazard (h) and cumulative hazard (H) is described as follows. The times of observed failure or censoring are put into ascending order, such that $t_{(1)} \leqslant t_{(2)} \leqslant \ldots t_{(n)}$. Each observation then receives its reverse rank if there are no ties. That is, the reverse rank of $t_{(n)}$, symbolized by r_n, is 1, that of $t_{(n-1)}$ is 2, and so on. The reverse ranks of $t_{(2)}$ and $t_{(1)}$ given by r_{n-1} and r_n respectively. The occurrence of tied observations in survival data is both a frequent and complicating feature. As the 'test' mentioned here is not a strict statistical tool but rather a visual technique which provides no exact significance level, little is lost by choosing midranks for tied data. The hazard for time $t_{(i)}$ is only calculated if the death occurred (i.e. $h_i = 0$ if the observation is censored). The hazard is given by the inverse of the reverse rank of the death:

$$h_i = 1/r_i = 1/(n-i+1).$$

The cumulative hazard is the sum of these:

$$H_i = \sum_{j=1}^{i} h_j.$$

(This description has referred to death as the event studied. For the duration of latency the event is the occurrence of cercarial release.) These relationships are summarized in Table 2, where an example with a specific censoring pattern is shown.

The relationship between the cumulative hazard and the time of event depends on the distribution that events are assumed to follow. For a given value of H_i, an

Table 2. *Calculation of the cumulative hazard function*

i	Censored = 0, death = 1	Ascending times	Reverse ranks	Hazard	Cumulative hazard
1	1	$t_{(1)}$	$r_1 = n$	$h_1 = 1/n$	$H_1 = h_1 = 1/n$
2	1	$t_{(2)}$	$r_2 = n-1$	$h_2 = 1/(n-1)$	$H_2 = h_1 + h_2 = 1/n + 1/(n-1)$
3	1	$t_{(3)}$	$r_3 = n-2$	$h_3 = 1/(n-2)$	$H_3 = h_1 + \ldots + h_3$
4	0	$t_{(4)}$	$r_4 = n-3$	$h_4 = 0$	$H_4 = h_1 + \ldots + h_4$
5	1	$t_{(5)}$	$r_5 = n-4$	$h_5 = 1/(n-4)$	$H_5 = h_1 + \ldots + h_5$
⋮	⋮	⋮	⋮	⋮	⋮
$n-3$	1	$t_{(n-3)}$	$r_{n-3} = 4$	$h_{n-3} = 1/4$	$H_{n-3} = h_1 + \ldots + h_{n-3}$
$n-2$	0	$t_{(n-2)}$	$r_{n-2} = 3$	$h_{n-2} = 0$	$H_{n-2} = h_1 + \ldots + h_{n-2}$
$n-1$	0	$t_{(n-1)}$	$r_{n-1} = 2$	$h_{n-1} = 0$	$H_{n-1} = h_1 + \ldots + h_{n-1}$
n	1	$t_{(n)}$	$r_n = 1$	$h_n = 1$	$H_n = h_1 + \ldots + h_n$.

associated time $x(H_i)$ can be obtained. The relationships between x and H are derived in Nelson's paper. For the exponential distribution one has

$$x(H_i) = \lambda^{-1} H_i \tag{4.1}$$

and for the normal distribution

$$x(H_i) = \mu + \sigma\phi^{-1}(1 - e^{-H_i}),$$

where ϕ^{-1} is the inverse of the standard normal cumulative distribution function. Thus, to check the validity of the exponential assumption H_i and $t_{(i)}$ are plotted. By (4.1) these should form a straight line through the origin. A check that $\hat{\lambda}$, calculated by previous methods is reasonable is made by noting that the slope of this line is λ^{-1}. Checking the normal assumption is a little more laborious. For each H_i one calculates $1 - e^{-H_i}$ and obtains the normal deviate associated with this value from tables. The plot of $t_{(i)}$ by $\phi^{-1}(1 - e^{-H_i})$ should be straight, and checks of $\hat{\mu}$ and $\hat{\sigma}$ obtained above can be likewise made from the empirical intercept and slope of the line respectively.

In practice, it is handy to check the exponential assumption of time to prepatent death with the normal assumption of duration of latency together. This is because prepatent deaths are 'failures' for the exponential test and 'censoreds' for the normal test. Snails becoming patent at $t_{(i)}$ are 'censoreds' for exponential tests and 'failures' for normal tests. Snails lost for other reasons are censoreds for both cases. In studying both assumptions the values of V_i are ordered to determine the $t_{(i)}$. The hazard and cumulative hazard are calculated for the exponential chart if $a_i(1 - b_i) = 1$ and for the normal chart if $b_i = 1$. Table 3 continues this study for the example of Section 2, and Figs. 1 and 2 display the graphs. In Fig. 1 a straight line through the origin seems adequate for all points except the last, where $(H_i, U_{(i)}) = (0{\cdot}8154, 36)$. Under experimental conditions one might query the validity of this observation: it may be an outlier. When a line is forced through the origin excluding this point the empirical equation is

$$U_{(i)} = 61{\cdot}4H_i,$$

and when this point is included the equation is

$$U_{(i)} = 56{\cdot}3H_i.$$

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

Table 3. *Calculation of cumulative hazards for the example in Table* 1

(1) Checking the exponential assumption of time to prepatent death, and the normal assumption of duration of latency.

		Exponential				Normal					
i	$V_{(i)}$	$a_i(1-b_i)$	r_i	h_i	H_i	b_i	r_i	h_i	H_i	$1-e^{-H_i}$	$\phi^{-1}(1-e^{-H_i})$
1	1	1	35	0·0286	0·0286	0	35	—	—	—	—
2	3	0	34	—	—	0	34	—	—	—	—
3	5	1	32½	0·0308	0·0593	0	32½	—	—	—	—
4	5	1	32½	0·0308	0·0901	0	32½	—	—	—	—
5	8	0	31	—	—	0	31	—	—	—	—
6	10	1	30	0·0333	0·1234	0	30	—	—	—	—
7	11	1	29	0·0345	0·1579	0	29	—	—	—	—
8	12	1	28	0·0357	0·1936	0	28	—	—	—	—
9	13	0	27	—	—	0	27	—	—	—	—
10	15	1	26	0·0385	0·2321	0	26	—	—	—	—
11	16	0	25	—	—	0	25	—	—	—	—
12	17	1	24	0·0417	0·2738	0	24	—	—	—	—
13	19	0	23	—	—	0	23	—	—	—	—
14	22	1	22	0·0455	0·3192	0	22	—	—	—	—
15	26	0	21	—	—	0	21	—	—	—	—
16	27	1	19½	0·0513	0·3705	0	19	—	—	—	—
17	27	1	19½	0·0513	0·4218	0	19	—	—	—	—
18	29	1	18	0·0556	0·4773	0	18	—	—	—	—
19	30	0	17	—	—	1	17	0·0588	0·0588	0·0571	−1·580
20	31	0	15½	—	—	1	15½	0·0645	0·1233	0·1160	−1·195
21	31	0	15½	—	—	1	15½	0·0645	0·1879	0·1713	−0·949
22	33	1	14	0·0714	0·5488	0	14	—	—	—	—
23	34	0	12½	—	—	1	12½	0·0800	0·2679	0·2350	−0·722
24	34	0	12½	—	—	1	12½	0·0800	0·3479	0·2938	−0·542
25	35	1	10	0·1000	0·6488	0	10	—	—	—	—
26	35	0	10	—	—	1	10	0·1000	0·4479	0·3610	−0·356
27	35	0	10	—	—	1	10	0·1000	0·5479	0·4218	−0·197
28	36	1	6	0·1667	0·8154	0	6	—	—	—	—
29	36	0	6	—	—	1	6	0·1667	0·7145	0·5106	0·026
30	36	0	6	—	—	1	6	0·1667	0·8812	0·5857	0·217
31	36	0	6	—	—	1	6	0·1667	1·0479	0·6493	0·384
32	36	0	6	—	—	1	6	0·1667	1·2145	0·7031	0·533
33	37	0	3	—	—	1	3	0·3333	1·5479	0·7873	0·797
34	39	0	2	—	—	1	2	0·5000	2·0479	0·8710	1·131
35	43	0	1	—	—	1	1	1·0000	3·0479	0·9525	1·670

(The slopes should be compared with $1/\hat{\lambda}_1 = 57{\cdot}5$ from the previous section.) In Fig. 2 there are no extreme observations and a linear equation is empirically obtained:

$$U_{(i)} = 35{\cdot}4 + 3{\cdot}52\phi^{-1}(1-e^{-H_i}).$$

(The intercept and slope should be compared with $\hat{\mu} = 35{\cdot}7$ and $\hat{\sigma} = 3{\cdot}21$ from the previous section This graphical approach is one way to obtain initial estimates for the Newton–Raphson technique, an alternative being to use the mean and variance estimates from the observed durations of patency.) It is hardly surprising to find a close agreement between the estimates obtained this way and the maximum likelihood values of Section 2. Nor is it surprising to find that straight

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

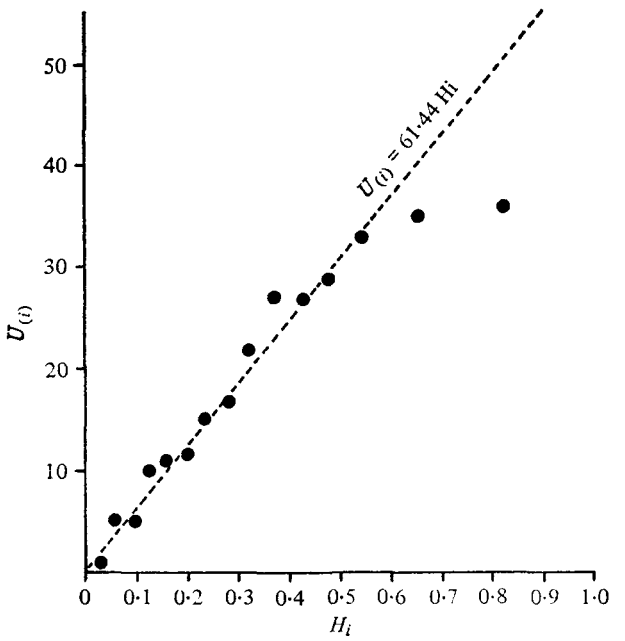

Fig. 1. Plot of cumulative hazard and event time to check exponential assumption of distribution of time to death of prepatent snails.

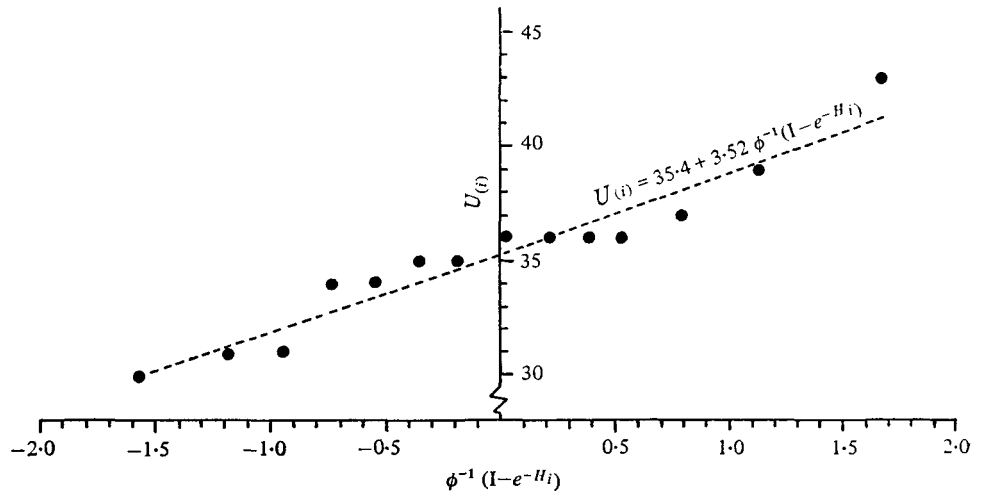

Fig. 2. Plot of $\phi^{-1}(1-e^{-H_i})$ and event time to check normal assumption of distribution of latency period.

lines seem reasonable as the data were generated from true exponential and normal distributions. (The peculiarity of the (0·8154, 36) point is surprising, however, but is not entirely unexpected from a random process!)

Testing the assumption that duration of life after patency is exponential is similarly performed. In this case the $t_i = U_i - V_i$ are ranked for those cases with $b_i = 1$. Cases with $a_i = 1$ are considered as 'failures' and those with $a_i = 0$ are

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

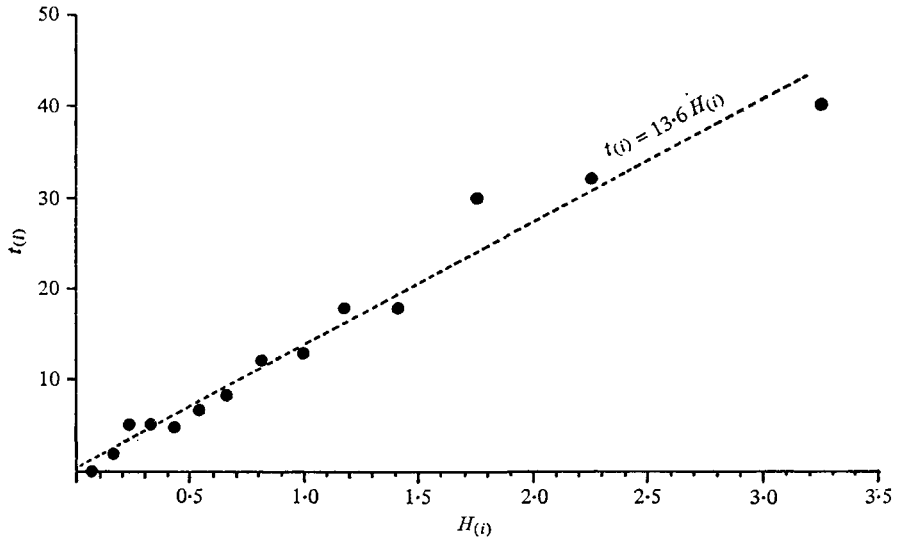

Fig. 3. Plot of cumulative hazard and event time to check exponential assumption of distribution of time to death of patent snails.

Table 4. *Calculation of cumulative hazards for the example in Table* 1

(2) Checking the exponential assumption of time to death following patency.

i	$t_{(i)}$	$a_i b_i$	r_i	h_i	H_i
1	0	1	14	0·0714	0·0714
2	2	1	13	0·0769	0·1484
3	5	1	11	0·0909	0·2393
4	5	1	11	0·0909	0·3302
5	5	1	11	0·0909	0·4211
6	7	1	9	0·1111	0·5322
7	8	1	8	0·1250	0·6572
8	12	1	7	0·1492	0·8000
9	13	1	6	0·1667	0·9667
10	18	1	$4\frac{1}{2}$	0·2222	1·1889
11	18	1	$4\frac{1}{2}$	0·2222	1·4112
12	30	1	3	0·3333	1·7445
13	32	1	2	0·5000	2·2445
14	40	1	1	1·0000	3·2445

'censoreds'. Table 4 and Fig. 3 display the relevant calculations and appropriate chart. No extreme observation is noted as in Fig. 1, the straight line is clear and the slope of a line forced through the origin is 13·6, which compares favourably with the value of $(1/\hat{\lambda}_3 =\ )13{\cdot}9$ from the second section.

Lastly, it is worth noting that probability paper to facilitate the graphing of the $t_{(i)}$ and H_i has been prepared. One supplier of this is TEAM, Box 25, Tamworth, N.H., USA 03886, as noted by Nelson.

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

5. EMBELLISHMENTS

Two further alterations in the model of Section 3 are discussed here. When there is evidence that the hazard plot in Section 4 is obviously not linear, the assumption of exponential survival must be relaxed. This is most readily done by assuming the time to death follows a Weibull distribution. The fitting of a Weibull distribution in place of an exponential distribution for the death of both prepatent and patent snail is first described in this section. Secondly, it is possible to alter the model to make allowance for factors possibly affecting the latent period or life-span, such as age at infection or water temperature. Statistically, this is a generalization of the model to allow for covariates. At first, allowance for covariates will be made in the primary model, following which it is easy to adapt the Weibull model as well.

When the assumption that the duration of life for an organism is exponentially distributed is invalid, two alternatives are usually considered, both of which require a second parameter to produce age-dependent hazard rates. In so far as the choice of distribution may be explained by an underlying process giving rise to a mathematical statement, then the Gamma distribution is an appealing generalization of the exponential. While heuristically this distribution seems advantageous, there is a considerable cost to pay both in analytic and computational terms. S. M. Gore in her work on extensions of the Lagakos model has investigated the use of the Gamma density in this setting and found it unacceptable for this application. A much more tractable generalization, suggested by Gore, is the Weibull distribution. For a discussion of the relative merits of these two distributions the reader is referred to chapter 4 of the book by Gross & Clark (1975).

The density function for the Weibull distribution is given by

$$f_j^*(t_i) = \theta_j \gamma_j t_i^{\gamma_j - 1} \exp(-\theta_j t_i^{\gamma_j}).$$

For $j = 1$ or 3 and $t_i \geqslant 0$, $i = 1, \ldots, n$. (The subscript $j = 2$ has been dropped for the reason given in Section 2.) It can be seen that the exponential distributions are recovered by fixing $\gamma_j = 1$ and $\theta_j = \lambda_j$ for $j = 1$ and 3. The log probability densities (excluding constants) of the four events corresponding to equations (3.1)–(3.4) follow:

$$\begin{aligned}
\log P(E_1) &= -\theta_1 U_i^{\gamma_1} + \log\{1 - F(\xi_i)\} \\
&= -\theta_1 V_i^{\gamma_1} + \log\{1 - F(\xi_i)\}, \\
\log P(E_2) &= -\theta_1 V_i^{\gamma_1} - \theta_3 (U_i - V_i)^{\gamma_3} - \log \sigma - \tfrac{1}{2}\xi_1^2, \\
\log P(E_3) &= -\theta_1 V_i^{\gamma_1} + \log \gamma_3 \theta_3 + (\gamma_3 - 1)\log(U_i - V_i) - \theta_3 (U_i - V_i)^{\gamma_3} - \log \sigma + \tfrac{1}{2}\xi_1^2, \\
\log P(E_4) &= \log \theta_1 \gamma_1 + (\gamma_1 - 1)\log U_i - \theta_1 U_i^{\gamma_1} + \log\{1 - F(\xi_i)\} \\
&= \log \theta_1 \gamma_1 + (\gamma_1 - 1)\log U_i - \theta_1 U_i^{\gamma_1} + \log\{1 - F(\xi_i)\}.
\end{aligned}$$

Hence the full log-likelihood excluding constants (corresponding to 3.5) is:

$$\begin{aligned}
L^* = &-\theta_1 \Sigma V_i^{\gamma_1} - \theta_3 \Sigma (U_i - V_i)^{\gamma_3} + \Sigma a_i b_i \{\log \theta_3 \gamma_3 + (\gamma_3 - 1)\log(U_i - V_i)\} \\
&+ \Sigma a_i (1 - b_i)\{\log \theta_1 \gamma_1 + (\gamma_1 - 1)\log U_i\} - \tfrac{1}{2}\Sigma[b_i(\log \sigma^2 + \xi_i^2) \\
&+ (1 - b_i)\log\{1 - F(\xi_i)\}].
\end{aligned}$$

(Note that $\log(U_i - V_i)$ is not undefined when $a_i b_i = 1$.)

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

The derivatives of L* with respect to μ and σ are the same as those of L. The derivatives with respect to the Weibull parameters follow. The estimation of parameters of Weibull distribution with censored observations is described by A. C. Cohen (1965). First derivatives:

$$\frac{\partial L^*}{\partial \theta_1} = -\Sigma V_i^{\gamma_1} + \Sigma \frac{a_i(1-b_i)}{\theta_1},$$

$$\frac{\partial L^*}{\partial \theta_3} = -\Sigma (U_i - V_i)^{\gamma_3} + \Sigma \frac{a_i b_i}{\theta_3},$$

$$\frac{\partial L^*}{\partial \gamma_1} = -\theta_1 \gamma_1 \Sigma V_i^{\gamma_1 - 1} + \Sigma a_i (1-b_i)\left(\frac{1}{\theta_1} + \log U_i\right),$$

$$\frac{\partial L^*}{\partial \gamma_3} = -\theta_3 \gamma_3 \Sigma (U_i - V_i)^{\gamma_3 - 1} + \Sigma a_i b_i \left\{\frac{1}{\theta_3} + \log (U_i - V_i)\right\}.$$

Second derivatives.

$$\frac{\partial^2 L^*}{\partial \theta_1^2} = -\Sigma \frac{a_i(1-b_i)}{\theta_1^2},$$

$$\frac{\partial^2 L^*}{\partial \theta_1 \partial \gamma_1} = -\gamma_1 \Sigma V_i^{\gamma_1 - 1},$$

$$\frac{\partial^2 L^*}{\partial \gamma_1^2} = -\theta_1 \Sigma V_i^{\gamma_1 - 2}\{V_i - \gamma_1(\gamma_1 - 1)\},$$

$$\frac{\partial^2 L^*}{\partial \theta_3^2} = -\Sigma \frac{a_i b_i}{\theta_3^2},$$

$$\frac{\partial^2 L^*}{\partial \theta_3 \partial \gamma_3} = -\gamma_3 \Sigma (U_i - V_i)^{\gamma_3 - 1},$$

$$\frac{\partial^2 L^*}{\partial \gamma_3^2} = -\theta_3 \Sigma (U_i - V_i)^{\lambda_3 - 2}\{(U_i - V_i) - \gamma_3(\gamma_3 - 1)\},$$

$$\frac{\partial^2 L^*}{\partial \theta_1 \partial \theta_3} = \frac{\partial^2 L^*}{\partial \theta_1 \partial \gamma_3} = \frac{\partial^2 L^*}{\partial \gamma_1 \partial \theta_3} = \frac{\partial^2 L^*}{\partial \gamma_1 \partial \gamma_3} = 0.$$

Unlike the exponential case, the maximum-likelihood extimates for all parameters now require iterative calculations. The matrix of second derivatives for all six parameters has canonical form and inversion is conveniently performed by inverting a series of three two-dimensional matrices.

A few words of warning are warranted here. The primary model involved four parameters and relaxing the exponential assumption for one or both survival times (or prepatent and patent snails) will require estimation of five or six parameters. It is not surprising to observe a much better fit when additional parameters are included in a formulation. There are approximate statistical measures to assess whether inclusion of a parameter produces a 'significantly' better fit: for example, Wilk's large sample likelihood ratio test (see Silvey, 1975,

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

chapter 7). The experimenter wishing to distinguish between models is recommended to decide on an empirical basis as well. To this end, both models might be fitted and estimated curves plotted. The estimated survival curves may have more immediate interpretation, but the distinction between models will be more pronounced if the age specific death rates or hazard functions are plotted. The decision is not exclusively statistical and should involve the experimenter's idea of what is a relevant difference.

It is possible, of course, to see if the Weibull provides an adequate fit to the data. Following the methods of Nelson (1972) described in Section 4, if H is the cumulative function and $x(H)$ the time of event dependent on H, one has

$$\log\{x(H_i)\} = \frac{1}{\gamma_j}\log H_i - \frac{1}{\gamma_j}\log\theta_j.$$

By plotting the log of the ranked time of event (either U_i or $U_i - V_i$) against the value of $\log H_i$ a straight line should result. Furthermore, from this one can obtain rough estimates of γ_j and θ_j.

Perhaps the duration of latency or the survival of snails depends on external factors such as the temperature of the water. There are two methods to investigate this possibility. The most straightforward approach would be to perform strictly controlled experiments (at a constant temperature) and obtain estimates (with variance estimates) of the parameters λ_1, λ_3, μ and σ. These four parameters can be estimated under a series of temperatures (or conditions) and their behaviour easily studied. An alternative is to make allowance for covariates or auxiliary variables in the formulation. Extensions along these lines are described in the original paper by Lagakos (1976, Section 3.2). In what follows, extensions of this form are made to all three stages whereas the experimenter may only wish to use one portion. (For example, only duration of latency and survival of patent snails may be affected by moderate temperature changes, while prepatent survival is not.)

Let us assume that s_j covariates are measured in process j on individual i and these are

$$(X_{j1i}, X_{j2i}, X_{j3i}, \ldots, X_{js_ji}).$$

(Here $j = 1$ for parameter λ_1, $j = 2$ for μ and σ and $j = 3$ for λ_3.) For the exponential processes, multiplicative hazard functions are assumed: i.e.

$$\lambda_{ji} = \exp(\beta_{j1}X_{j1i} + \beta_{j2}X_{j2i} + \ldots + \beta_{js_j}X_{js_ji}) \quad (j = 1 \text{ or } 3,\ i = 1, \ldots, n).$$

The former model is recovered by assuming $s_j = 1$ and $X_{j1i} = 1$ for all i and both j's. The covariates are related to μ in the duration of latency by:

$$\mu_i = \beta_{21}X_{21i} + \beta_{22}X_{22i} + \ldots + \beta_{2s_2}X_{2s_2i} \quad \text{for} \quad i = 1, \ldots, n.$$

Thus we seek estimates of the β_{jk} for $j = 1, 2, 3$ and $k = 1, \ldots s_j$ which maximize (3.5). The derivatives (3.6) to (3.14) are easily adapted to this end by noting that

$$\frac{\partial\lambda_{ji}}{\partial\beta_{jk}} = X_{jki}\lambda_{ji} \quad \text{for} \quad i = 1, \ldots n;\ j = 1 \text{ or } 3;\ k = 1, \ldots, s_j$$

and

$$\frac{\partial\mu_i}{\partial\beta_{2k}} = X_{2ki} \quad \text{for} \quad k = 1, \ldots, s_2.$$

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

and using the chain rule. Hence one has

$$\frac{\partial L}{\partial \beta_{1k}} = \Sigma a_i(1-b_i)\, X_{1ki} - \Sigma V_i X_{1ki}\lambda_{1i},$$

$$\frac{\partial L}{\partial \beta_{2k}} = \left[\frac{1}{\sigma}\Sigma\{b_i\xi_i + (1-b_i)Z_i\}X_{2ki}\right],$$

$$\frac{\partial L}{\partial \beta_{3k}} = \Sigma a_i b_i - \Sigma(U_i - V_i)X_{3ki}\lambda_{3i}.$$

The second derivatives are:

$$\frac{\partial^2 L}{\partial \beta_{1k}\partial \beta_{1l}} = -\Sigma V_i \lambda_{1i} X_{1ki} X_{1li} \quad \text{for} \quad k = 1, \ldots, s_1 \text{ and } l = 1, \ldots, s_1,$$

$$\frac{\partial^2 L}{\partial \beta_{2k}\partial \beta_{2l}} = -\frac{1}{\sigma^2}\Sigma\{b_i + (1-b_i)A_i\}X_{2ki}X_{2li} \quad \text{for} \quad k = 1, \ldots, s_2 \text{ and} \quad l = 1 \ldots s_2,$$

$$\frac{\partial^2 L}{\partial \beta_{2k}\partial \sigma} = \frac{1}{\sigma^2}\Sigma\{2b_i\xi_i + (1-b_i)B_i\}X_{2ki} \quad \text{for} \quad k = 1, \ldots s_2,$$

$$\frac{\partial^2 L}{\partial \beta_{3k}\partial \beta_{3l}} = -\Sigma(U_i - V_i)\lambda_{3i}X_{3ki}X_{3li} \quad \text{for} \quad k = 1, \ldots, s_3 \quad \text{and} \quad l = 1, \ldots, s_3.$$

(The functions A_i and B_i were given in 3.15 and 3.16.)

It is becoming clear that inclusion of several covariates may be intuitively appealing but entails considerably greater computational complexity. None of the paramaters β_{jk} can be obtained explicitly as λ_1 and λ_3 could, and iterative procedures are necessary.

It remains to merge the two embellishments described in this section and make allowance for covariates in the Weibull model. This is very similar to those considerations for the exponential. Let

$$\theta_{ji} = \exp(\beta_{j1}X_{j1i} + \beta_{j2}X_{j2i} + \ldots + \beta_{js_j}X_{js_ji}) \quad \text{for} \quad j = 1 \text{ or } 3, i = 1 \ldots n$$

and note that

$$\frac{\partial \theta_{ji}}{\partial \beta_{jk}} = X_{jki}\theta_{ji}.$$

The derivatives of L^* for determining the β_{jk} for $j = 1$ or 3 and $k = 1, \ldots, s_j$ follow.

$$\frac{\partial L^*}{\partial \beta_{1k}} = \Sigma\{a_i(1-b_i) - \theta_{1i}V_i^{\gamma_1}\}X_{1ki} \quad (k = 1, \ldots, s_1 \quad i = 1, \ldots, n),$$

$$\frac{\partial L^*}{\partial \beta_{3k}} = \Sigma\{a_i b_i - (U_i - V_i)^{\gamma_3}\}X_{3ki} \quad (k = 1, \ldots, s_3 \quad i = 1, \ldots, n),$$

https://doi.org/10.1017/S0022172400025845 Published online by Cambridge University Press

with

$$\frac{\partial^2 L^*}{\partial \beta_{1k}\beta_{1l}} = -\Sigma\theta_{1i}V_i^{\gamma_1}X_{1ki}X_{1li} \quad (k, l = 1, \ldots, s_i \quad i = 1, \ldots, n),$$

$$\frac{\partial^2 L^*}{\partial \beta_{1k}\partial\gamma_1} = -\gamma_1\Sigma\theta_{1i}V_i^{\gamma_1-1}X_{1ki} \quad (k = 1, \ldots, s_1 \quad i = 1, \ldots n),$$

$$\frac{\partial^2 L^*}{\partial \beta_{3k}\partial\beta_{3l}} = -\Sigma\theta_{3i}(U_i - V_i)^{\gamma_3}X_{3ki}X_{3li} \quad (k, l = 1, \ldots, s_3 \quad i = 1, \ldots, n),$$

$$\frac{\partial^2 L^*}{\partial \beta_{3k}\partial\gamma_3} = -\gamma_3\Sigma\theta_{3i}(U_i - V_i)^{\gamma_3-1}X_{3ki} \quad (k = 1 \ldots, s_3 \quad i = 1, \ldots, n),$$

and

$$\frac{\partial^2 L^*}{\partial \beta_{1k}\partial\beta_{3l}} = \frac{\partial^2 L^*}{\partial \beta_{1k}\partial\gamma_3} = \frac{\partial^2 L^*}{\partial \beta_{3l}\partial\gamma_1} = \frac{\partial^2 L^*}{\partial \gamma_1\partial\gamma_3} = 0 \quad \text{for } k = 1, \ldots, s_1 \quad \text{and} \quad l = 1, \ldots, s_3.$$

6. SUMMARY

By means of techniques of analyses of survival data developed for cancer trials it is possible to study aspects of the natural history of the infection of schistosomiasis on the intermediate host of transmission, the snail.

The simultaneous study of three response variables is largely based on a model of Lagakos (1976). When using this approach in the schistosomiasis setting it seems inappropriate to assume that one process, the duration of latency, follows an exponential distribution. Thus this stage is modified to follow a normal distribution and the derivatives required to obtain maximum-likelihood estimates and approximate variances of all parameters are provided.

Simple graphical tools for assessing the validity of distributional assumptions in survival data are available from industrial research. The reader's attention is drawn to a paper by Nelson (1972). The relevance and application of these methods to the current problem are described in Section 4.

In the event that the times to death of prepatent and patent snails do not follow exponential distributions as assumed in the primary model, a further modification is introduced to enable either or both to follow Weibull densities.

Lastly it is possible to adapt both the primary model of Section Three and the modified model of Section Five to allow for the inclusion of auxiliary variables or covariates. Again the required derivatives to obtain maximum likelihood estimates and approximate variances are provided.

I wish to thank Sheila Gore, Stuart Pocock and Professor Michael Healy for many useful discussions. Sheila Gore and Professor Peter Armitage provided many recommendations for the improvement of the manuscript for which I am most grateful.

REFERENCES

Cohen, A. C. (1963). Progressively censored samples in life testing. *Technometrics* **5**, 327–39.

Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics* **7**, 579–88.

Cohen, J. E. (1973). Selective host mortality in a catalytic model applied to schistosomiasis. *American Naturalist* **107**, 199–212.

Cohen, J. E. (1977). Mathematical models of schistosomiasis. *Annual Review of Ecology and Systematics* **8**, 209–33.

Gross, A. J. & Clark, V. A. (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley.

Harter, H. L. & Moore, A. M. (1965). Maximum-likelihood estimation of the parameters of Gamma and Weibull populations from complete and from censored samples. *Technometrics* **7**, 639–43.

Jordan, P. & Webbe, G. (1969). *Human Schistosomiasis*. London: Heinemann.

Lagakos, S. W. (1976). A stochastic model for censored survival data in the presence of an auxiliary variable. *Biometrics* **32**, 551–9.

Muench, H. (1959). *Catalytic Models in Epidemiology*, Cambridge: Harvard University Press.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945–66.

Silvey, S. D. (1975). *Statistical Inference*. London: Chapman and Hall.

Sturrock, R. F. & Webbe, G. (1971). The application of catalytic models to schistosomiasis in snails. *Journal of Helminthology* **45**, 189–200.

Sturrock, R. F., Cohen, J. E. & Webbe, G. (1975). Catalytic curve analysis of schistosomiasis in snails. *Annals of Tropical Medicine and Parasitology* **69**, 133–4.