Regression Models and Life-Tables

By D. R. Cox

Imperial College, London

[Read before the Royal Statistical Society, at a meeting organized by the Research Section, on Wednesday, March 8th, 1972, Mr M. J. R. Healy in the Chair]

Summary

The analysis of censored failure times is considered. It is assumed that on each individual are available values of one or more explanatory variables. The hazard function (age-specific failure rate) is taken to be a function of the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time. A conditional likelihood is obtained, leading to inferences about the unknown regression coefficients. Some generalizations are outlined.

Keywords: LIFE TABLE; HAZARD FUNCTION; AGE-SPECIFIC FAILURE RATE; PRODUCT LIMIT ESTIMATE; REGRESSION; CONDITIONAL INFERENCE; ASYMPTOTIC THEORY; CENSORED DATA; TWO-SAMPLE RANK TESTS; MEDICAL APPLICATIONS; RELIABILITY THEORY; ACCELERATED LIFE TESTS.

1. Introduction

Life tables are one of the oldest statistical techniques and are extensively used by medical statisticians and by actuaries. Yet relatively little has been written about their more formal statistical theory. Kaplan and Meier (1958) gave a comprehensive review of earlier work and many new results. Chiang in a series of papers has, in particular, explored the connection with birth–death processes; see, for example, Chiang (1968). The present paper is largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables and more generally to the incorporation of regression-like arguments into life-table analysis. The arguments are asymptotic but are relevant to situations where the sampling fluctuations are large enough to be of practical importance. In other words, the applications are more likely to be in industrial reliability studies and in medical statistics than in actuarial science. The procedures proposed are, especially for the two-sample problem, closely related to procedures for combining contingency tables; see Mantel and Haenzel (1959), Mantel (1963) and, especially for the application to life tables, Mantel (1966). There is also a strong connection with a paper read recently to the Society by R. and J. Peto (1972).

We consider a population of individuals; for each individual we observe either the time to "failure" or the time to "loss" or censoring. That is, for the censored individuals we know only that the time to failure is greater than the censoring time.

Denote by T a random variable representing failure time; it may be discrete or continuous. Let $\mathscr{F}(t)$ be the survivor function,

$$\mathscr{F}(t) = \text{pr}(T \geqslant t)$$

and let $\lambda(t)$ be the hazard or age-specific failure rate. That is,

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{\text{pr}(t \leqslant T < t+\Delta t \,|\, t \leqslant T)}{\Delta t}. \tag{1}$$

Note that if T is discrete, then

$$\lambda(t) = \sum \lambda_{u_j} \delta(t-u_j), \tag{2}$$

where $\delta(t)$ denotes the Dirac delta function and $\lambda_t = \mathrm{pr}(T = t \mid T \geqslant t)$. By the product law of probability $\mathscr{F}(t)$ is given by the product integral

$$\mathscr{F}(t) = \prod_{u=0}^{t-0} \{1-\lambda(u)\,du\} = \lim \prod_{k=0}^{r-1} \{1-\lambda(\tau_k)(\tau_{k+1}-\tau_k)\}, \tag{3}$$

the limit being taken as all $\tau_{k+1}-\tau_k$ tend to zero with $0 = \tau_0 < \tau_1 < \ldots < \tau_{r-1} < \tau_r = t$. If $\lambda(t)$ is integrable this is

$$\exp\left\{-\int_0^t \lambda(u)\,du\right\}, \tag{4}$$

whereas if $\lambda(t)$ is given by (2), the product integral is

$$\prod_{u_j<t} (1-\lambda_{u_j}). \tag{5}$$

If the distribution has both discrete and continuous components the product integral is a product of factors (4) and (5).

2. The Product-limit Method

Suppose observations are available on n_0 independent individuals and, to begin with, that the failure times are identically distributed in the form specified in Section 1. Let n individuals be observed to failure and the rest be censored. The rather strong assumption will be made throughout that the only information available about the failure time of a censored individual is that it exceeds the censoring time. This assumption is testable only if suitable supplementary information is available. Denote the distinct failure times by

$$t_{(1)} < t_{(2)} < \ldots < t_{(k)}. \tag{6}$$

Further let $m_{(i)}$ be the number of failure times equal to $t_{(i)}$, the multiplicity of $t_{(i)}$; of course $\sum m_{(i)} = n$, and in the continuous case $k = n$, $m_{(i)} = 1$.

The set of individuals at risk at time $t-0$ is called the *risk set* at time t and denoted $\mathscr{R}(t)$; this consists of those individuals whose failure or censoring time is at least t. Let $r_{(i)}$ be the number of such individuals for $t = t_{(i)}$. The product-limit estimate of the underlying distribution is obtained by taking estimated conditional probabilities that agree exactly with the observed conditional frequencies. That is,

$$\hat{\lambda}(t) = \sum_{i=1}^{k} \frac{m_{(i)}}{r_{(i)}} \delta(t-t_{(i)}). \tag{7}$$

Correspondingly,

$$\hat{\mathscr{F}}(t) = \prod_{u=0}^{t-0} \{1-\hat{\lambda}(u)\,du\} = \prod_{t_{(i)}<t} \left\{1-\frac{m_{(i)}}{r_{(i)}}\right\}. \tag{8}$$

For uncensored data this is the usual sample survivor function; some of the asymptotic properties of (8) are given by Kaplan and Meier (1958) and by Efron (1967) and can be used to adapt to the censored case tests based on sample cumulative distribution function.

The functions (7) and (8) are maximum-likelihood estimates in the family of all possible distributions (Kaplan and Meier, 1958). However, as in the uncensored case,

this property is of limited importance and the best justification is essentially (7). The estimates probably also have a Bayesian interpretation involving a very "irregular" prior.

If the class of distributions is restricted, either parametrically or by some such condition as requiring $\lambda(t)$ to be monotonic or smooth, the maximum-likelihood estimates will be changed. For the monotone hazard case with uncensored data, see Grenander (1956). The smoothing of estimated hazard functions has been considered by Watson and Leadbetter (1964a, b) for the uncensored case.

3. Regression Models

Suppose now that on each individual one or more further measurements are available, say on variables $z_1, \ldots, z_p$. We deal first with the notationally simpler case when the failure-times are continuously distributed and the possibility of ties can be ignored. For the jth individual let the values of $\mathbf{z}$ be $\mathbf{z}_j = (z_{1j}, \ldots, z_{pj})$. The z's may be functions of time. The main problem considered in this paper is that of assessing the relation between the distribution of failure time and $\mathbf{z}$. This will be done in terms of a model in which the hazard is

$$\lambda(t; \mathbf{z}) = \exp(\mathbf{z}\boldsymbol{\beta})\,\lambda_0(t), \tag{9}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\lambda_0(t)$ is an unknown function giving the hazard function for the standard set of conditions $\mathbf{z} = \mathbf{0}$. In fact $(\mathbf{z}\boldsymbol{\beta})$ can be replaced by any known function $h(\mathbf{z}, \boldsymbol{\beta})$, but this extra generality is not needed at this stage. The following examples illustrate just a few possibilities.

Example 1. *Two-sample problem.* Suppose that there is just one z variable, $p = 1$, and that this takes values 0 and 1, being an indicator variable for the two samples. Then according to (9) the hazards in samples 0 and 1 are respectively $\lambda_0(t)$ and $\psi\lambda_0(t)$, where $\psi = e^{\beta}$. In the continuous case the survivor functions are related (Lehmann, 1953) by $\mathscr{F}_1(t) = \{\mathscr{F}_0(t)\}^{\psi}$. There is an obvious extension for the k sample problem.

Example 2. *The two-sample problem; extended treatment.* We can deal with more complicated relationships between the two samples than are contemplated in Example 1 by introducing additional time-dependent components into $\mathbf{z}$. Thus if $z_2 = tz_1$, where z_1 is the binary variable of Example 1, the hazard in the second sample is

$$\psi e^{\beta_2 t}\,\lambda_0(t). \tag{10}$$

Of course in defining z_2, t could be replaced by any known function of t; further, several new variables could be introduced involving different functions of t. This provides one way of examining consistency with a simple model of proportional hazards. In fitting the model and often also in interpretation it is convenient to reparametrize (10) in the form

$$\rho \exp\{\beta_2(t - t^*)\}, \tag{11}$$

where t^* is any convenient constant time somewhere near the overall mean. This will avoid the more extreme non-orthogonalities of fitting. All the points connected with this example extend to the comparison of several samples.

Example 3. *Two-sample problem with covariate.* By introducing into the models of Examples 1 and 2 one or more further z variables representing concomitant variables, it is possible to examine the relation between two samples adjusting for the presence of concomitant variables.

Example 4. *Regression.* The connection between failure-time and regressor variables can be explored in an obvious way. Note especially that by introducing functions of t, effects other than constant multiplication of the hazard can be included.

4. Analysis of Regression Models

There are several approaches to the analysis of the above models. The simplest is to assume $\lambda_0(t)$ constant, i.e. to assume an underlying exponential distribution; see, for example, Chernoff (1962) for some models of this type in the context of accelerated life tests. The next simplest is to take a two-parameter family of hazard functions, such as the power law associated with the Weibull distribution or the exponential of a linear function of t. Then standard methods such as maximum likelihood can be used; to be rigorous extension of the usual conditions for maximum-likelihood formulae and theory would be involved to cover censoring, but there is little doubt that some such justification could be given. This is in many ways the most natural approach but will not be explored further in the present paper. In this approach a computationally desirable feature is that both probability density and survivor function are fairly easily found. A simple form for the hazard is not by itself particularly advantageous, and models other than (9) may be more natural. For a normal theory maximum-likelihood analysis of factorial experiments with censored observations, see Sampford and Taylor (1959), and for the parametric analysis of response times in bioassay, see, Sampford (1954).

Alternatively we may restrict $\lambda_0(t)$ qualitatively, for example by assuming it to be monotonic or to be a step function (a suggestion of Professor J. W. Tukey). The latter possibility is related to a simple spline approximation to the log survivor function.

In the present paper we shall, however, concentrate on exploring the consequence of allowing $\lambda_0(t)$ to be arbitrary, main interest being in the regression parameters. That is, we require our method of analysis to have sensible properties whatever the form of the nuisance function $\lambda_0(t)$. Now this is a severe requirement and unnecessary in the sense that an assumption of some smoothness in the distribution $\mathscr{F}_0(t)$ would be reasonable. The situation is parallel to that arising in simpler problems when a nuisance parameter is regarded as completely unknown. It seems plausible in the present case that the loss of information about $\boldsymbol{\beta}$ arising from leaving $\lambda_0(t)$ arbitrary is usually slight; if this is indeed so the procedure discussed here is justifiable as a reasonably cautious approach to the study of $\boldsymbol{\beta}$. A major outstanding problem is the analysis of the relative efficiency of inferences about $\boldsymbol{\beta}$ under various assumptions about $\lambda_0(t)$.

The general attitude taken is that parametrization of the dependence on $\mathbf{z}$ is required so that our conclusions about that dependence are expressed concisely; of course any form taken is provisional and needs examination in the light of the data. So far as the secondary features of the system are concerned, however, it is sensible to make a minimum of assumptions leading to a convenient analysis, provided that no major loss of efficiency is involved.

5. A Conditional Likelihood

Suppose then that $\lambda_0(t)$ is arbitrary. No information can be contributed about $\boldsymbol{\beta}$ by time intervals in which no failures occur because the component $\lambda_0(t)$ might conceivably be identically zero in such intervals. We therefore argue conditionally on the set $\{t_{(i)}\}$ of instants at which failures occur; in discrete time we shall condition

also on the observed multiplicities $\{m_{(i)}\}$. Once we require a method of analysis holding for all $\lambda_0(t)$, consideration of this conditional distribution seems inevitable.

For the particular failure at time $t_{(i)}$, conditionally on the risk set $\mathscr{R}(t_{(i)})$, the probability that the failure is on the individual as observed is

$$\exp\{\mathbf{z}_{(i)}\boldsymbol{\beta}\}\Big/\sum_{l\in\mathscr{R}(t_{(i)})}\exp\{\mathbf{z}_{(l)}\boldsymbol{\beta}\}. \tag{12}$$

Each failure contributes a factor of this nature and hence the required conditional log likelihood is

$$L(\boldsymbol{\beta})=\sum_{i=1}^{k}\mathbf{z}_{(i)}\boldsymbol{\beta}-\sum_{i=1}^{k}\log\Big[\sum_{l\in\mathscr{R}(t_{(i)})}\exp\{\mathbf{z}_{(l)}\boldsymbol{\beta}\}\Big]. \tag{13}$$

Direct calculation from (13) gives for $\xi,\eta=1,\ldots,p$

$$U_\xi(\boldsymbol{\beta})=\frac{\partial L(\boldsymbol{\beta})}{\partial\beta_\xi}=\sum_{i=1}^{k}\{z_{(\xi i)}-A_{(\xi i)}(\boldsymbol{\beta})\}, \tag{14}$$

where

$$A_{(\xi i)}(\boldsymbol{\beta})=\frac{\sum z_{\xi l}\exp(\mathbf{z}_l\boldsymbol{\beta})}{\sum\exp(\mathbf{z}_l\boldsymbol{\beta})}, \tag{15}$$

the sum being over $l\in\mathscr{R}(t_{(i)})$. That is, $A_{(\xi i)}(\boldsymbol{\beta})$ is the average of z_ξ over the finite population $\mathscr{R}(t_{(i)})$, using an "exponentially weighted" form of sampling. Similarly

$$\mathscr{I}_{\xi\eta}(\boldsymbol{\beta})=-\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\beta_\xi\,\partial\beta_\eta}=\sum_{i=1}^{k}C_{(\xi\eta i)}(\boldsymbol{\beta}), \tag{16}$$

where

$$C_{(\xi\eta i)}(\boldsymbol{\beta})=\{\sum z_{\xi l}z_{\eta l}\exp(\mathbf{z}_l\boldsymbol{\beta})/\sum\exp(\mathbf{z}_l\boldsymbol{\beta})\}-A_{(\xi i)}(\boldsymbol{\beta})\,A_{(\eta i)}(\boldsymbol{\beta}) \tag{17}$$

is the covariance of z_ξ and z_η in this form of weighted sampling.

To calculate the expected value of (16) it would be necessary to know the times at which individuals who failed would have been censored had they not failed. This information would often not be available and in any case might well be thought irrelevant; this point is connected with difficulties of conditionality at the basis of a sampling theory approach to statistics (Pratt, 1962). Here we shall use asymptotic arguments in which (16) can be used directly for the estimation of variances, $\boldsymbol{\beta}$ being replaced by a suitable estimate. For a rigorous justification, assumptions about the censoring times generalizing those of Breslow (1970) would be required. It would not be satisfactory to assume that the censoring times are random variables distributed independently of the $\mathbf{z}$'s. For instance in the two-sample problem censoring might be much more severe in one sample than in the other.

Maximum-likelihood estimates of $\boldsymbol{\beta}$ can be obtained by iterative use of (14) and (16) in the usual way. Significance tests about subsets of parameters can be derived in various ways, for example by comparison of the maximum log likelihoods achieved. Relatively simple results can, however, be obtained for testing the global null hypothesis, $\boldsymbol{\beta}=\mathbf{0}$. For this we treat $\mathbf{U}(\mathbf{0})$ as asymptotically normal with zero mean vector and with covariance matrix $\mathscr{I}(\mathbf{0})$. That is, the statistic

$$\{\mathbf{U}(\mathbf{0})\}^{\mathrm{T}}\{\mathscr{I}(\mathbf{0})\}^{-1}\{\mathbf{U}(\mathbf{0})\} \tag{18}$$

has, under the null hypothesis, an asymptotic chi-squared distribution with p degrees of freedom.

We have from (14) and (15) that

$$\mathbf{U}_\xi(\mathbf{0}) = \sum_{i=1}^{k} (z_{(\xi i)} - A_{(\xi i)}), \tag{19}$$

where $A_{(\xi i)} = A_{(\xi i)}(\mathbf{0})$ is the mean of z_ξ over $\mathscr{R}(t_{(i)})$. Further, from (16),

$$\mathscr{I}_{\xi\eta}(\mathbf{0}) = \sum_{i=1}^{k} C_{(\xi\eta i)}, \tag{20}$$

where $C_{(\xi\eta i)} = C_{(\xi\eta i)}(\mathbf{0})$ is the covariance of z_ξ and z_η in the finite population $\mathscr{R}(t_{(i)})$. The form of weighted sampling associated with general $\boldsymbol{\beta}$ has reduced to random sampling without replacement.

6. Analysis in Discrete Time

Unfortunately it is quite likely in applications that the data will be recorded in a form involving ties. If these are small in number a relatively *ad hoc* modification of the above procedures will be satisfactory. To cover the possibility of an appreciable number of ties, we generalize (9) formally to discrete time by

$$\frac{\lambda(t; \mathbf{z})\, dt}{1 - \lambda(t; \mathbf{z})\, dt} = \exp(\mathbf{z}\boldsymbol{\beta}) \frac{\lambda_0(t)\, dt}{1 - \lambda_0(t)\, dt}. \tag{21}$$

In the continuous case this reduces to (9); in discrete time $\lambda(t; \mathbf{z})\, dt$ is a non-zero probability and (21) is a logistic model.

The typical contribution (12) to the likelihood now becomes

$$\exp\{\mathbf{s}_{(i)}\boldsymbol{\beta}\} \Big/ \sum_{l \in \mathscr{R}(t_{(i)}; m_{(i)})} \exp\{\mathbf{s}_{(l)}\boldsymbol{\beta}\}, \tag{22}$$

where $\mathbf{s}_{(i)}$ is the sum of $\mathbf{z}$ over the individuals failing at $t_{(i)}$ and the notation in the denominator means that the sum is taken over all distinct sets of $m_{(i)}$ individuals drawn from $\mathscr{R}(t_{(i)})$.

Thus the full conditional log likelihood is

$$\sum_{i=1}^{k} \mathbf{s}_{(i)}\boldsymbol{\beta} - \sum_{i=1}^{k} \log \left[\sum_{l \in \mathscr{R}(t_{(i)}; m_{(i)})} \exp\{\mathbf{s}_{(l)}\boldsymbol{\beta}\} \right].$$

The derivatives can be calculated as before. In particular,

$$U_\xi(\mathbf{0}) = \sum_{i=1}^{k} \{s_{(\xi i)} - m_{(i)} A_{(\xi i)}\}, \tag{23}$$

$$\mathscr{I}_{\xi\eta}(\mathbf{0}) = \sum_{i=1}^{k} \frac{m_{(i)}\{r_{(i)} - m_{(i)}\}}{\{r_{(i)} - 1\}} C_{(\xi\eta i)}. \tag{24}$$

Note that (24) gives the exact covariance matrix when the observations $z_{(\xi i)}$ and the totals $s_{(\xi i)}$ are drawn randomly without replacement from the *fixed* finite populations $\mathscr{R}(t_{(1)}), \ldots, \mathscr{R}(t_{(k)})$. In fact, however, the population at one time is influenced by the outcomes of the "trials" at previous times.

7. The Two-sample Problem

As an illustration, consider the two-sample problem with the proportional hazard model of Section 3, Example 1. Here $p = 1$ and we omit the first suffix on the indicator variable. Then

$$U(0) = n_1 - \sum_{i=1}^{k} m_{(i)} A_{(i)}, \tag{25}$$

$$\mathscr{I}(0) = \sum_{i=1}^{k} \frac{m_{(i)}\{r_{(i)} - m_{(i)}\}}{\{r_{(i)} - 1\}} A_{(i)}\{1 - A_{(i)}\}, \tag{26}$$

where $A_{(i)}$ is the proportion of the risk population $\mathscr{R}(t_{(i)})$ that have $z = 1$, i.e. belong to sample 1, and n_1 is the total number of failures in sample 1. An asymptotic two-sample test is thus obtained by treating

$$U(0)/\sqrt{\mathscr{I}(0)} \tag{27}$$

as having a standard normal distribution under the null hypothesis. This is different from the procedure of Gehan who adapted the Wilcoxon test to censored data (Gehan, 1965; Efron, 1967; Breslow, 1970). The test has been considered in some detail by Peto and Peto (1972).

The test (27) is formally identical with that obtained by setting up at each failure point a 2×2 contingency table (sample 1, sample 2) (failed, survived). To test for the presence of a difference between the two samples the information from the separate tables can then be combined (Cochran, 1954; Mantel and Haenzel, 1959; Mantel, 1963). The application of this to life tables is discussed especially by Mantel (1966). Note, however, that whereas the test in the contingency table situation is, at least in principle, exact, the test here is only asymptotic, because of the difficulties associated with specification of the stopping rule. Formally the same test was given by Cox (1959) for a different life-table problem where there is a single sample with two types of failure and the hypothesis under test concerns the proportionality of the hazard function for the two types.

When there is a non-zero value of β, the "weighted" average of a single observation from the risk population $\mathscr{R}(t_{(i)})$ is

$$A_{(i)}(\beta) = \frac{e^{\beta} A_{(i)}}{1 - A_{(i)} + e^{\beta} A_{(i)}} \tag{28}$$

and the maximum-likelihood equation $U(\hat{\beta}) = 0$ gives, when all failure times are distinct,

$$\sum_{i=1}^{k} \frac{e^{\hat{\beta}} A_{(i)}}{1 - A_{(i)} + e^{\hat{\beta}} A_{(i)}} = n_1. \tag{29}$$

If $\hat{\beta}$ is thought to be close to some known constant, it may be useful to linearize (29). In particular, if $\hat{\beta}$ is small, we have as an approximation to the maximum-likelihood estimate

$$\hat{\beta}_0 = (n_1 - \sum A_{(i)}) / \sum A_{(i)}\{1 - A_{(i)}\}.$$

The procedures of this section involve only the ranked data, i.e. are unaffected by an arbitrary monotonic transformation of the time scale. Indeed the same is true for any of the results in Section 4 provided that the $\mathbf{z}$'s are not functions of time. While

the connection with the theory of rank tests will not be explored in detail, it is worth examining the form of the test (27) for uncensored data with all failure times distinct. For this, let the failure times in sample 1 have ranks $c_1 < c_2 < \ldots < c_{n_1}$ in the ranking of the full data. At the ith largest observed failure time, individuals with ranks $n, n-1, \ldots, i$ are at risk, so that

$$A_{(i)} = \frac{1}{n-i+1} \sum_{l=1}^{n_1} H(c_l - i), \tag{30}$$

where $H(x)$ is the unit Heaviside function,

$$H(x) = \begin{cases} 1 & (x \geqslant 0), \\ 0 & (x < 0). \end{cases} \tag{31}$$

Thus, by (25),

$$U(0) = n_1 - \sum_{l=1}^{n_1} \sum_{i=1}^{c_l} \frac{1}{n-i+1}$$

$$= n_1 - \sum_{l=1}^{n} e_{nc_l}, \tag{32}$$

where e's are the expected values of the order statistics in a random sample of size n from a unit exponential distribution. The test based on (32) is asymptotically fully efficient for the comparison of two exponential distributions (Savage, 1956; Cox, 1964). Further, by (26),

$$\mathscr{I}(0) = \sum_{l=1}^{n_1} e_{nc_l} - \sum_{l=1}^{n_1} (1 + 2n_1 - 2l) v_{nc_l}, \tag{33}$$

where

$$v_{nc_l} = \sum_{i=1}^{c_l} \frac{1}{(n-i+1)^2} \tag{34}$$

is the variance of an exponential order statistic.

Here the test statistic is, under the null hypothesis, a constant minus the total of a random sample of size n_1 drawn without replacement from the finite population $\{e_{n1}, \ldots, e_{nn}\}$. The exact distribution can in principle be obtained and in particular it can be shown that

$$E\{U(0)\} = 0, \quad \text{var}\{U(0)\} = \frac{n_1(n-n_1)(n-e_{nn})}{n(n-1)}. \tag{35}$$

There is not much point in this case in using the more complicated asymptotic formula (33), especially as fairly simple more refined approximations to the distribution of the test statistic are available (Cox, 1964). It can easily be verified that

$$E\{\mathscr{I}(0)\} \sim \text{var}\{U(0)\}. \tag{36}$$

8. Estimation of Distribution of Failure-time

Once we have obtained the maximum-likelihood estimate of β, we can consider the estimation of the distribution associated with the hazard (10) either for $\mathbf{z} = \mathbf{0}$, or for some other given value of $\mathbf{z}$. Thus to estimate $\lambda_0(t)$ we need to generalize (7).

To do this we take $\lambda_0(t)$ to be identically zero, except at the points where failures have occurred, and carry out a separate maximum-likelihood estimation at each such failure point. For the latter it is convenient to write the contribution to $\lambda_0(t)$ at $t_{(i)}$ in the form

$$\frac{\pi_{(i)}\exp(-\boldsymbol{\beta}\tilde{\mathbf{z}}_{(i)})}{1-\pi_{(i)}+\pi_{(i)}\exp(-\boldsymbol{\beta}\tilde{\mathbf{z}}_{(i)})}\delta(t-t_{(i)}),$$

where $\tilde{\mathbf{z}}_{(i)}$ is an arbitrary constant to be chosen; it is useful to take $\tilde{\mathbf{z}}_{(i)}$ as approximately the mean in the relevant risk set. The maximum-likelihood estimate of $\pi_{(i)}$ can then be shown to satisfy

$$\hat{\pi}_{(i)}=\frac{m_{(i)}}{r_{(i)}}-\frac{\hat{\pi}_{(i)}(1-\hat{\pi}_{(i)})}{r_{(i)}}\sum_{j\in R(t_{(i)})}\frac{\exp\{\hat{\boldsymbol{\beta}}(\mathbf{z}_j-\tilde{\mathbf{z}}_{(i)})\}-1}{1-\hat{\pi}_{(i)}+\hat{\pi}_{(i)}\exp\{\hat{\boldsymbol{\beta}}(\mathbf{z}_j-\tilde{\mathbf{z}}_{(i)})\}}, \tag{37}$$

which can be solved by iteration. The suggested choice of $\tilde{\mathbf{z}}_{(i)}$ is designed to make the second term in (37) small. Note that in the single-sample case, the second term is identically zero. Once (37) is solved for all i, we have by the product integral formula

$$\hat{\mathscr{F}}_0(t)=\prod_{t_{(i)}<t}\left\{1-\frac{\hat{\pi}_{(i)}\exp(-\hat{\boldsymbol{\beta}}\tilde{\mathbf{z}}_{(i)})}{1-\hat{\pi}_{(i)}+\hat{\pi}_{(i)}\exp(-\hat{\boldsymbol{\beta}}\tilde{\mathbf{z}}_{(i)})}\right\}. \tag{38}$$

For an estimate at a given non-zero $\mathbf{z}$, replace $\exp(-\hat{\boldsymbol{\beta}}\tilde{\mathbf{z}}_{(i)})$ by $\exp\{\hat{\boldsymbol{\beta}}(\mathbf{z}-\tilde{\mathbf{z}}_{(i)})\}$. Alternative simpler procedures would be worth having (Mantel, 1966).

9. Bivariate Life Tables

We now consider briefly the extension of life-table arguments to multivariate data. Suppose for simplicity that there are two types of failure time for each individual represented by random variables T_1 and T_2. For instance, these might be the failure-times of two different but associated components; observations may be censored on neither, one or both components. For analogous problems in bioassay, see Sampford (1952).

The joint distribution can be described in terms of hazard functions $\lambda_{10}(t)$, $\lambda_{20}(t)$, $\lambda_{21}(t\,|\,u)$, $\lambda_{12}(t\,|\,u)$, where

$$\left.\begin{aligned}\lambda_{p0}(t)&=\lim_{\Delta t\to 0+}\frac{\mathrm{pr}(t\leqslant T_p<t+\Delta t\,|\,t\leqslant T_1,t\leqslant T_2)}{\Delta t}\quad(p=1,2),\\ \lambda_{21}(t\,|\,u)&=\lim_{\Delta t\to 0+}\frac{\mathrm{pr}(t\leqslant T_2<t+\Delta t\,|\,t\leqslant T_2,T_1=u)}{\Delta t}\quad(u<t),\end{aligned}\right\} \tag{39}$$

with a similar definition for $\lambda_{12}(t\,|\,u)$. It is easily shown that the bivariate probability density function $f(t_1,t_2)$ is given by

$$f(t_1,t_2)=\exp\left[-\int_0^{t_1-0}\{\lambda_{10}(u)+\lambda_{20}(u)\}\,du-\int_{t_1+0}^{t_2-0}\lambda_{21}(u\,|\,t_1)\,du\right]\lambda_{10}(t_1)\,\lambda_{21}(t_2\,|\,t_1), \tag{40}$$

for $t_2\geqslant t_1$, with again an analogous expression for $t_2\leqslant t_1$. It is fairly easy to show formally that a necessary and sufficient condition for the independence of T_1 and T_2 is

$$\lambda_{12}(t\,|\,u)=\lambda_{10}(t),\quad \lambda_{21}(t\,|\,u)=\lambda_{20}(t), \tag{41}$$

as is obvious on general grounds. Note also that if $\mathscr{F}(t_1, t_2)$ is the joint survivor function

$$\lambda_{10}(t) = -\frac{1}{\mathscr{F}(t,t)}\left[\frac{\partial \mathscr{F}(t,u)}{\partial t}\right]_{u=t}, \quad \lambda_{12}(t|u) = -\frac{\partial^2 \mathscr{F}(t,u)}{\partial t\, \partial u}\Big/\frac{\partial \mathscr{F}(t,u)}{\partial u}. \tag{42}$$

Dependence on further variables $\mathbf{z}$ can be indicated in the same way as for (11). The simplest model would have the same function of $\mathbf{z}$ multiplying all four hazard functions, although this restriction is not essential.

Estimation and testing would in principle proceed as before, although grouping of the conditioning u variable seems necessary in the parts of the analysis concerning the function $\lambda_{12}(t|u)$ and $\lambda_{21}(t|u)$.

Further generalizations which will not, however, be explored here are to problems in multidimensional time and to problems connected with point processes (Cox and Lewis, 1972; Cox, 1972).

10. An Example

To illustrate some of the above results, it is convenient to take data of Freireich *et al.* used by Gehan (1965) and several subsequent authors. Table 1 gives the ordered times for two samples of individuals; censored values are denoted with asterisks. Table 2 outlines the calculation of the simple test statistic $U(0)$ and its asymptotic variance. The failure instants and their multiplicities $m_{(i)}$ are listed; $A_{(i)}$ is the proportion of the relevant risk population in sample 1.

Table 1

Times of remission (weeks) of leukemia patients
(Gehan, 1965, from Freireich *et al.*)

Sample 0 (drug 6-MP)	6*, 6, 6, 6, 7, 9*, 10*, 10, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*
Sample 1 (control)	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

* Censored.

The value of $U(0) = n_1 - \sum m_{(i)} A_{(i)}$ is 10·25 with an asymptotic standard error $\sqrt{\mathscr{I}(0)}$ of 2·50. The critical ratio of just over 4 compares with about 3·6 for the generalized Wilcoxon test of Gehan (1965). The overwhelming significance of the difference is in line with one's qualitative impression of the data.

The technique used to find $\hat{\beta}$ was direct computation of the log likelihood as a function of β and of a further parameter γ to be described in a moment. This, while not the best way of getting maximum-likelihood estimates on their own, is useful in enabling various approximate tests and confidence regions to be found in a unified manner.

To examine possible departures from the simple model of proportional hazards, the procedure of Example 2 of Section 3 was followed, taking as in (11) the hazard in sample 1 to be a time-dependent multiple of that in sample 0 of the form

$$\exp\{\beta + \gamma(t-10)\}\lambda_0(t); \tag{43}$$

the arbitrary constant 10 is inserted to achieve approximate orthogonality of estimation of the two parameters, being chosen as a convenient value in the centre of the range.

A test of the global null hypothesis $\beta = \gamma = 0$ could be done via the test statistic (20) but is not very relevant here. Instead the log likelihood (15) was computed

TABLE 2

Main quantities for the test of the null hypothesis for the data of Table 1

"Failure" time		*Risk population*			*Multiplicity*	
Sample 0	*Sample* 1	*No. in sample* 0	*No. in sample* 1	$r_{(i)}$	$A_{(i)}$	$m_{(i)}$
23	23	6	1	7	0·1429	2
22	22	7	2	9	0·2222	2
	17	10	3	13	0·2308	1
16		11	3	14	0·2143	1
	15	11	4	15	0·2667	1
13		12	4	16	0·2500	1
	12, 12	12	6	18	0·3333	2
	11, 11	13	8	21	0·3810	2
10		15	8	23	0·3478	1
	8, 8, 8, 8	16	12	28	0·4286	4
7		17	12	29	0·4138	1
6, 6, 6		21	12	33	0·3636	3
	5, 5	21	14	35	0·4000	2
	4, 4	21	16	37	0·4324	2
	3	21	17	38	0·4474	1
	2, 2	21	19	40	0·4750	2
	1, 1	21	21	42	0·5000	2

$$U(0) = n_1 - \sum m_{(i)} A_{(i)} = 10{\cdot}25;$$

$$\mathscr{I}(0) = \sum \frac{m_{(i)}\{r_{(i)} - m_{(i)}\}}{r_{(i)} - 1} A_{(i)}\{1 - A_{(i)}\} = 6.2570.$$

directly for a grid of points in the (β, γ) plane. Note that in (15) the first term is $21\beta - 28\gamma$; for instance, the coefficient -28 is the sum of the values $(t-10)$ over the individuals in sample 1. The logarithmic second term is simple for those time points at which there is a single completed time, $m_{(i)} = 1$; for example corresponding to the time 7 there is a term in the log likelihood

$$-\log(17 + 12e^{\beta - 3\gamma}),$$

the risk set at this time consisting of 17 individuals from sample 0 and 12 from sample 1. For points of higher multiplicity, the situation is more complicated, because all possible samples of size $m_{(i)}$ from the risk population have to be considered; fortunately all the samples have the same totals of the two relevant variables. For example, for the point 6, of multiplicity 3, we have to consider the total of all samples of size 3 drawn from the relevant risk population and this leads to a term

$$-\log\left\{\binom{21}{3} + \binom{21}{2}\binom{12}{1} e^{\beta - 4\gamma} + \binom{21}{1}\binom{12}{2} e^{2\beta - 8\gamma} + \binom{12}{3} e^{3\beta - 12\gamma}\right\}. \qquad (44)$$

To avoid unduly large numbers, it might often be convenient to divide each term in the logarithm by a suitable constant, but this was not done in the present case.

The maximum-likelihood estimate of β when $\gamma = 0$ is $\hat{\beta} = 1{\cdot}65$. Thus the ratio of the hazards is estimated as $e^{\hat{\beta}} = 5{\cdot}21$; if the distributions were exponential, this would be the ratio of means. Confidence limits for β, subject to $\gamma = 0$, can be obtained either by computing the second derivative $\mathscr{I}(\hat{\beta})$ or directly from the log likelihood. With the latter method, approximate 95 per cent confidence limits for β of $(0{\cdot}78, 2{\cdot}60)$ are obtained from those values for which the log likelihood is within $\frac{1}{2} \times 1{\cdot}96^2 = 1{\cdot}92$ of its maximum value. An alternative test of the null hypothesis $\beta = 0$ is obtained by comparing the log likelihood at $\beta = 0$ and $\beta = \hat{\beta}$; the difference of 7·43 corresponds to chi-squared of 14·9 and hence to a standardized deviate of 3·86, in reasonable agreement with test based on $U(0)$.

The inclusion of the extra parameter γ provides a test of the adequacy of the assumption of simply related hazards. In fact the additional log likelihood achieved by the extra parameter, about 0·01, is small, even suspiciously small. Confidence limits for γ are, at the 95 per cent level, approximately $-0{\cdot}12$ and 0·14. Thus any marked departure from the proportional hazard model is not likely to be a smooth monotonic change with t. Further details of the likelihood function will not be given here. It is, however, quadratic to a close approximation and the particular parametrization chosen achieved almost exact orthogonality.

Finally, we consider graphical techniques, which are likely to be particularly useful for data more extensive than the present set. A first step is to obtain unconditional estimates of the separate survivor functions by (8). For sample 1 this gives the ordinary sample survivor function, there being no censoring. For sample 0, we get the product limit estimate. Now consider estimation of the survivor functions under the model of proportional hazards; the constrained maximum-likelihood estimates of the survivor functions in the two samples are given by (37) and (38). Iterative solution of the 17 equations of the form (37) took in all $\frac{1}{2}$ sec. on the CDC 6600; $\tilde{z}$ was chosen separately for each risk set so that $e^{\hat{\beta}\tilde{z}}$ equalled the mean of $e^{\hat{\beta}z}$ over the risk set in question.

Fig. 1 shows the four estimated functions. Discrepancy with the model of proportional hazards would be shown by clear departures of the conditional from the unconstrained survivor curves. More elaborate versions of this analysis are certainly possible, in which, for instance, plots are made on a non-linear scale, or in which residuals from the constrained fit are formed, or in which the analysis is presented in tabulated rather than graphical form. The graphical analysis confirms the consistency of the data with a model of proportional hazards.

Only a very brief note will be added here about alternative approaches to the analysis. If exponential distributions are assumed the relevant statistics are the total periods at risk, namely 359 weeks and 182 weeks, and the total numbers of failures 9 and 22 respectively. Approximate 95 per cent confidence limits for the log ratio of means can be obtained via the F distribution with (18, 44) degrees of freedom. They are 0·83 and 2·43, as compared with 0·78 and 2·60 from the earlier analysis.

An analysis with a step function for $\lambda_0(.)$ is barely feasible with the limited amount of data available. The procedure is to divide the time scale into cells, for instance 0–10 weeks and 11–20 weeks. Numbers of failures and periods at risk are calculated for each cell and hence ratios of rates derived. Provided they are consistent for th$_e$

different cells the ratios can then be combined into a single summary statistic with confidence limits. In the present example this approach does not lead to essentially different conclusions.

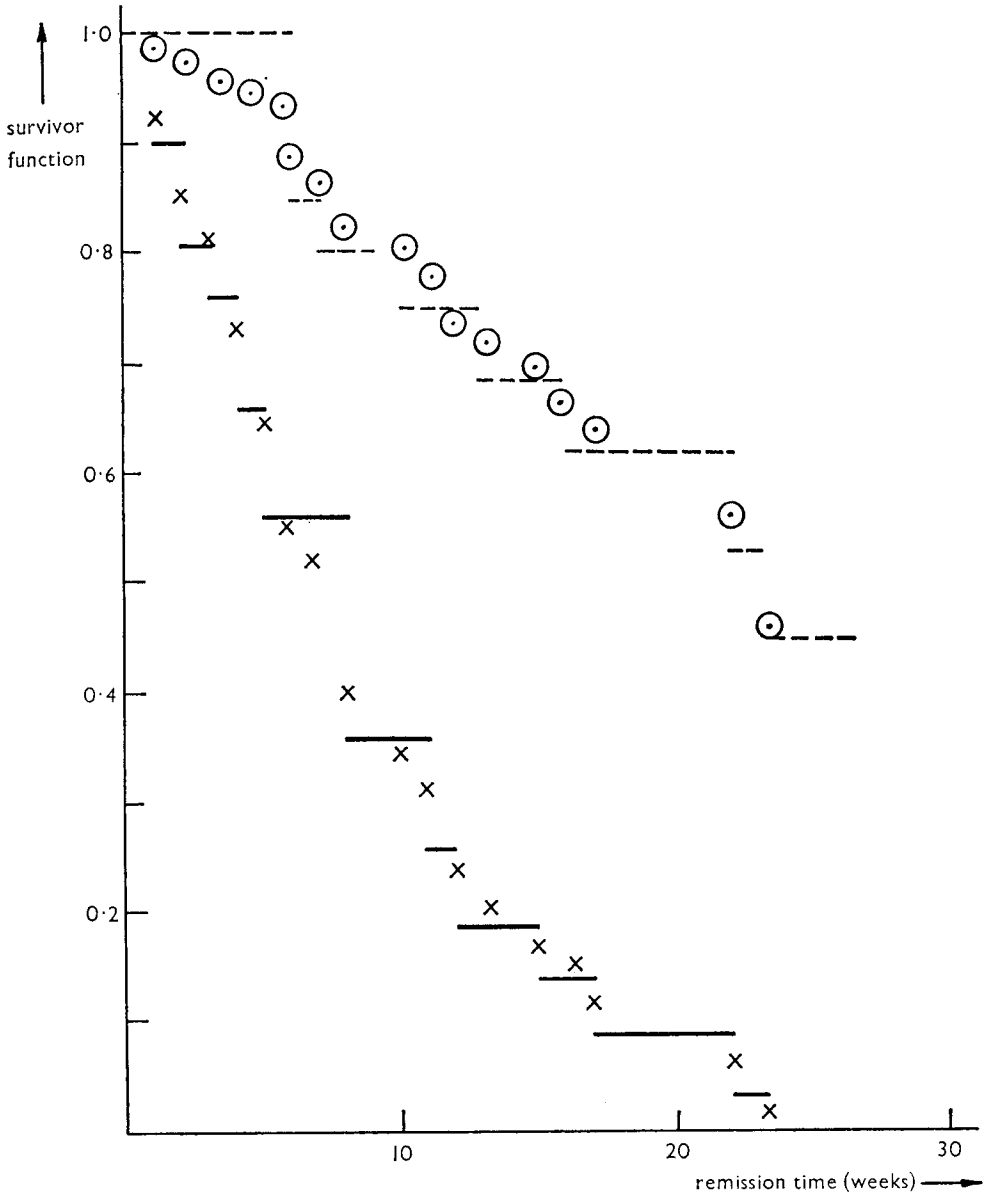

FIG. 1. Empirical survivor functions for data of Table 1. Product limit estimate, – – – – – –, sample 0 (6-MP); ———, sample 1 (control). Estimate constrained by proportionality: ⊙, sample 0; ×, sample 1. For clarity, the constrained estimates are indicated by the left ends of the defining horizontal lines.

A third possibility is the use of the Weibull distribution. If we assume a common index in the two samples we may fit by maximum-likelihood distribution functions in the form

$$1-\exp\{-(\rho x/\kappa)^{\nu}\}, \quad 1-\exp\{-(\kappa\rho x)^{\nu}\}.$$

The maximum-likelihood estimate of the index is $\hat{\nu} = 1{\cdot}3$ and the maximized log-likelihoods show that this is just significantly different from $\nu = 1{\cdot}0$ at the 5 per cent level. The explanation of the departure probably lies largely in the deficiency of small failure times in sample 0. Fitting of different indexes for the two samples has not been attempted. Approximate 95 per cent confidence limits for the log ratio of means can be derived in the usual way from the maximized log likelihoods and are 0·71 and 2·10; the maximum-likelihood estimate is $\log(\hat{\kappa}^2) = 1{\cdot}31$.

The data have been analysed in some detail to illustrate a number of relevant points. Many applications are likely to be more complicated partly because of larger sample sizes and partly because of the presence of a number of explanatory variables.

11. Physical Interpretation of Model

The model (9), which is the basis of this paper, is intended as a representation of the behaviour of failure-time that is convenient, flexible and yet entirely empirical. One of the referees has, however, suggested adding some discussion of the physical meaning of the model and in particular of its possible relevance to accelerated life testing. Suppose in fact that there is a variable s, called "stress", and that life tests are carried out at various levels of s. For simplicity we suppose that s is one-dimensional and that each individual is tested at a fixed level of s. The usual idea is that we are really interested in some standard stress, say $s = 1$, and which to use other values of s to get quick laboratory results as a substitute for a predictor of the expensive results of user trials.

Now in order that the distribution of failure-time at one level of stress should be related to that at some other level, the relationship being stable under a wide range of conditions, it seems necessary that the basic physical process of failure should be common at the different stress levels; and this is likely to happen only when there is a single predominant mode of failure. One difficulty of the problem is that of knowing enough about the physical process to be able to define a stress variable, i.e. a set of test conditions, with the right properties.

One of the simplest models proposed for the effect of stress on the distribution of failure-time is to assume that the mechanism of failure is identical at the various levels of s but takes place on a time-scale that depends on s. Thus if $\mathscr{F}(t; s)$ denotes the survivor function at stress s, this model implies that

$$\mathscr{F}(t; s) = \mathscr{F}\{g(s)\,t; 1\}, \tag{45}$$

where $g(s)$ is some function of s with $g(1) = 1$. Thus the hazard function at stress s is

$$g(s)\,\lambda_0\{g(s)\,t\}, \tag{46}$$

where $\lambda_0(.)$ is the hazard at $s = 1$. In particular if $g(s) = s^{\beta}$ and if $z = \log s$ this gives

$$e^{\beta z}\,\lambda_0(e^{\beta z}\,t). \tag{47}$$

This is similar to but different from the model (9) of this paper. A special set of conditions where (47) applies is where the individual is subject to a stream of shocks of randomly varying magnitudes until the cumulative shock exceeds some time-independent tolerance. If, for instance, all aspects of the process except the rate of incidence of shocks are independent of s, then (45) will apply.

If, however, the shocks are non-cumulative and failure occurs when a rather high threshold is first exceeded, failures occur in a Poisson process with a rate depending on s. A special model of this kind often used for thermal stress is to suppose that failure corresponds to the excedence of the activation energy of some process; then by the theory of rate processes (47) can be used with $\lambda_0(.) = 1$ and z equal to the reciprocal of absolute temperature.

As a quite different model suppose that some process of ageing goes on independently of stress. Suppose further that the conditional probability of failure at any time is the product of an instantaneous time-dependent term arising from the ageing process and a stress-dependent term; the model is non-cumulative. Then the hazard is

$$h(s)\,\lambda_0(t), \tag{48}$$

where $h(s)$ is some function of stress. Again if $h(s) = s^\beta$, the model becomes

$$e^{\beta z}\,\lambda_0(t) \tag{49}$$

exactly that of (9), where again $\lambda_0(t)$ is the hazard function at $s = 1$, $z = 0$. One special example of this model is rather similar to that suggested for (46), except that the critical tolerance varies in a fixed way with time and the shocks are non-cumulative, the rate of incidence of shocks depending on s. For another possibility, see Shooman (1968).

If hazard or survivor functions are available at various levels of s we might attempt an empirical discrimination between (46) and (48). Note, however, that if we have a Weibull distribution at $s = 1$, $\lambda_0(.)$ is a power function and (46) and (48) are identical. Then the models cannot be discriminated from failure-time distributions alone. That is, if we did want to make such a discrimination we must look for situations in which the distributions are far from the Weibull form. Of course the models outlined here can be made much more specific by introducing explicit stochastic processes or physical models. The wide variety of possibilities serves to emphasize the difficulty of inferring an underlying mechanism indirectly from failure times alone rather than from direct study of the controlling physical processes.

As a basis for rather empirical data reduction (9), possibly with time-dependent exponent, seems flexible and satisfactory.

Acknowledgements

I am grateful to the referees for helpful comments and to Professor P. Armitage, Mr P. Fisk, Dr N. Mantel, Professors J. W. Tukey and M. Zelen for references and constructive suggestions.

References

Breslow, N. (1970). A generalized Kruskal–Wallis test for comparing samples subject to unequal patterns of censoring. *Biometrika*, **57**, 579–594.

Chernoff, H. (1962). Optimal accelerated life designs for estimation. *Technometrics*, **4**, 381–408.

Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. New York: Wiley.

COCHRAN, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, **10**, 417–451.
COX, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure. *J. R. Statist. Soc.* B, **21**, 411–421.
—— (1964). Some applications of exponential ordered scores. *J. R. Statist. Soc.* B, **26**, 103–110.
—— (1972). The statistical analysis of dependencies in point processes. In *Symposium on Point Processes* (P. A. W. Lewis, ed.). New York: Wiley (to appear).
COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. *Proc. 6th Berkeley Symp.* (to appear).
EFRON, B. (1967). The two sample problem with censored data. *Proc. 5th Berkeley Symp.*, **4**, 831–853.
GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored samples. *Biometrika*, **52**, 203–224.
GRENANDER, U. (1956). On the theory of mortality measurement, I and II. *Skand. Akt.*, **39**, 90–96, 125–153.
KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.*, **53**, 457–481.
LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.*, **24**, 23–43.
MANTEL, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel–Haenzel procedure. *J. Am. Statist. Assoc.*, **58**, 690–700.
—— (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**, 163–170.
MANTEL, N. and HAENZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.*, **22**, 719–748.
PETO, R. and PETO, J. (1972). Asymptotically efficient rank invariant test procedures. *J. R. Statist. Soc.* A **135**, 185–206.
PRATT, J. W. (1962). Contribution to discussion of paper by A. Birnbaum. *J. Am. Statist. Assoc.*, **57**, 314–316.
SAMPFORD, M. R. (1952). The estimation of response-time distributions, II: Multi-stimulus distributions. *Biometrics*, **8**, 307–369.
—— (1954). The estimation of response-time distribution, III: Truncation and survival. *Biometrics*, **10**, 531–561.
SAMPFORD, M. R. and TAYLOR, J. (1959). Censored observations in randomized block experiments. *J. R. Statist. Soc.* B, **21**, 214–237.
SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics—the two-sample case. *Ann. Math. Statist.*, **27**, 590–615.
SHOOMAN, M. L. (1968). Reliability physics models. *IEEE Trans. on Reliability*, **17**, 14–20.
WATSON, G. S. and LEADBETTER, M. R. (1964a). Hazard analysis, I. *Biometrika*, **51**, 175–184.
WATSON, G. S. and LEADBETTER, M. R. (1964b). Hazard analysis, II. *Sankhyā*, A, **26**, 101–116.

DISCUSSION ON PROFESSOR COX'S PAPER

Professor F. DOWNTON (University of Birmingham): Professor Cox has given us a paper which is characteristically both elegant and useful. One can only regret that it is probably true that, as he says, "the applications are more likely to be in industrial reliability studies and in medical statistics than in actuarial science". Benjamin (1972) gave one reason for this when he said that to insurance companies the estimation of future mortality was the least of their problems; the major parameter in life insurance has become the interest rate on invested money. It would appear that insurance companies are, in general, extremely reluctant to take on special short-term risks, where the methods of this paper could be applied. One would have thought, however, that these methods could be used in non-life insurance. Would it be too outrageous to suggest that the recent failures in motor insurance would not have occurred if the companies concerned had read, applied and drawn the correct conclusions from this paper?

However, I do not wish to discuss practical applications, but to suggest that by giving his paper a somewhat restrictive title Professor Cox has been too modest. He has said that he does not wish to explore the connection of this paper with the theory of rank tests, so I hope he will forgive me if I do. Basically the approach adopted here is a mixture