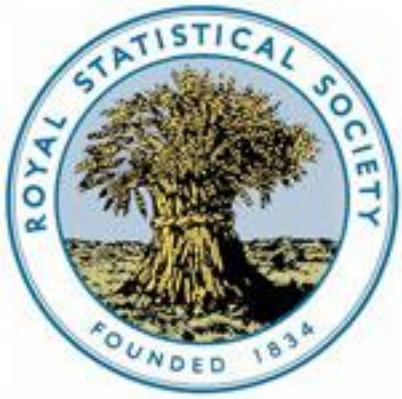

Regression Models and Life-Tables
Author(s): D. R. Cox
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 2 (1972), pp. 187-220
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2985181
Accessed: 19-04-2016 20:48 UTC

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

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

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),

$$\begin{aligned} U(0) &= n_1 - \sum_{l=1}^{n_1} \sum_{i=1}^{c_l} \frac{1}{n-i+1} \\ &= n_1 - \sum_{l=1}^{n} e_{nc_l}, \end{aligned} \tag{32}$$

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

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

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

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

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.

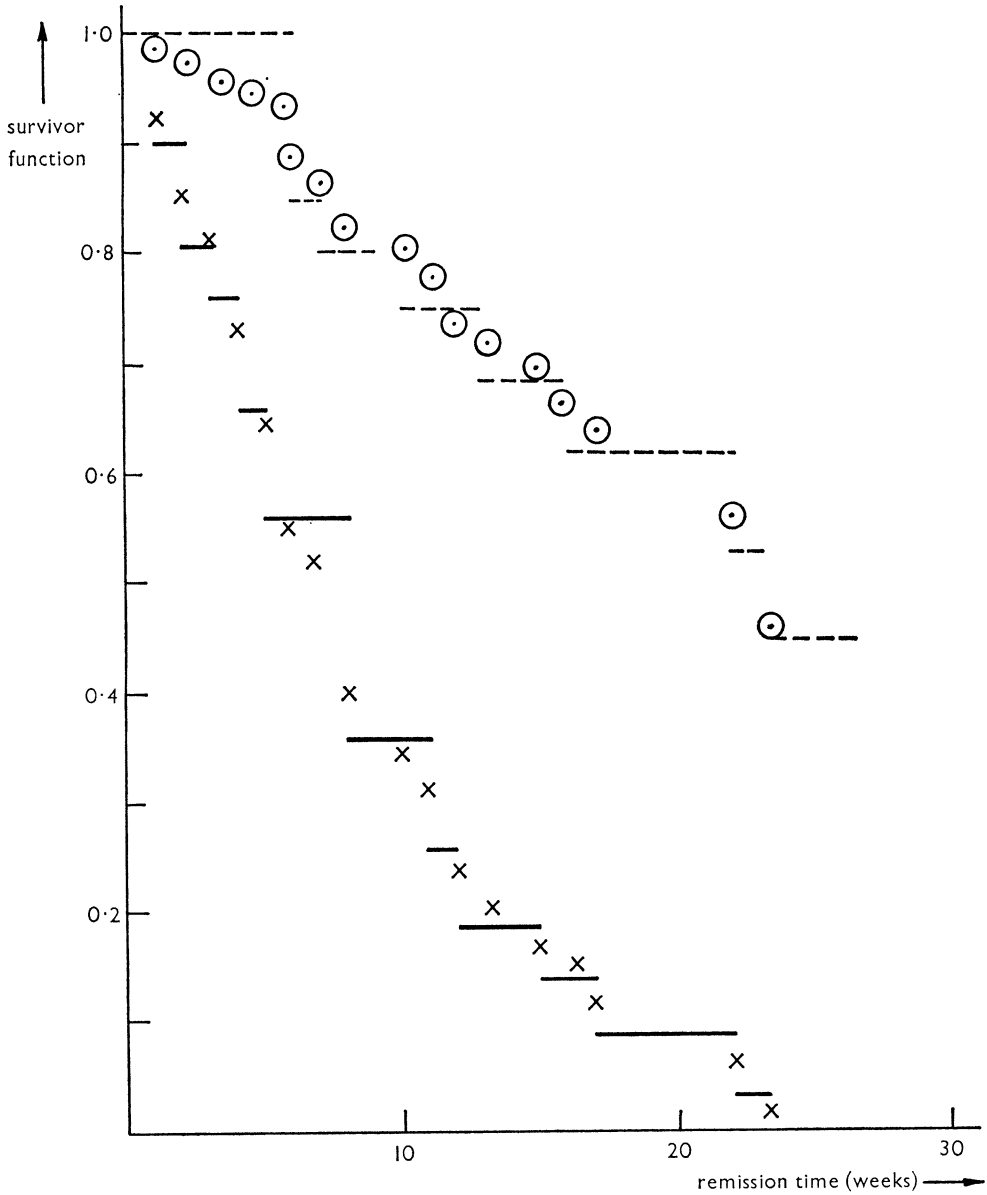

FIG. 1. Empirical survivor functions for data of Table 1. Product limit estimate, – – – – –, sample 0 (6-MP); ———, sample 1 (control). Estimate constrained by proportionality: ⊙, sample 0; ×, sample 1. For clarity, the constrained estimates are indicated by the left ends of the defining horizontal lines.

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

of the parametric and the non-parametric but this approach may be used to derive non-parametric test procedures of a more traditional kind. I will illustrate this for one class of problems.

The clue lies in his remark in Example 1 of Section 3 that for the two sample problem the basic model implies that we are concerned with a Lehmann-type family of distributions. This was also the condition found by Armitage (1959) for using the Dixon and Mood "sign test". It seems natural to ask first how a paired comparison design would respond to the treatment of this paper. We assume therefore that out of n pairs of results, r specimens given treatment A "failed" before their paired specimens, which had been given treatment B. For the remaining $n-r$ pairs the position was reversed. Then if the failure rates were $\lambda_0(t)$ and $\lambda_0(t)\,e^{\beta}$ for A and B, respectively, using the conditional argument of Section 5 of the paper, the probability, at the first failure time t_i of the ith pair, that failure occurred to the actual individual observed is $e^{z\beta}/(1+e^{\beta})$, where $z = 0$ or 1, according as the failure was of the specimen given treatment A or B, respectively. The log likelihood is then

$$L(\beta) = r\beta - n\log(1+e^{\beta}),$$

whence

$$U(\beta) = \frac{\partial L(\beta)}{\partial \beta} = r - \frac{ne^{\beta}}{1+e^{\beta}},$$

$$\frac{\partial^2 L(\beta)}{\partial \beta^2} = -\frac{ne^{\beta}}{(1+e\beta)^2} = E\left\{\frac{\partial^2 L(\beta)}{\partial \beta^2}\right\}.$$

Thus to test the hypothesis $\beta = 0$ we have the test statistic

$$4(r-n/2)^2/n,$$

whose distribution (if $\beta = 0$) is, asymptotically, χ^2 with one degree of freedom.

This is, of course, the "sign test" for the median and is a trivial result. However, paired comparisons are a special case of the randomized block design, and generalizing the method above yields test statistics for that situation different from those usually used.

We will assume that we have n blocks each containing the results for $p+1$ treatments, these results being ranked in order of preference in each block. Equivalently we may say that for each block we have an observation consisting of a permutation of the numbers 0 to p, representing the treatments arranged in order of preference.

We suppose that in the jth block the distributions underlying the ranking of the $p+1$ treatments are of the form

$$1-F_{i,j}(t) = \{1-F_j(t)\}^{k_i}, \quad i = 0, 1, \ldots, p; \quad j = 1, 2, \ldots, n.$$

The "standard" treatment corresponding to $i = 0$ may be chosen arbitrarily and we assume $k_0 = 1$. This distributional assumption is equivalent, in Professor Cox's terms, to a hazard function for the ith treatment in the jth block of the form

$$\lambda_j(t)\,e^{\beta_i}, \quad \text{with } \beta_i = \log_e k_i \quad (\beta_0 = 0).$$

We now need to use a slight generalization of the conditional argument of Section 5 to see that if we index the possible $(p+1)!$ permutations of treatments by $r = 1, 2, \ldots, (p+1)!$ then the conditional probability of obtaining the rth permutation may be written

$$\exp\left\{\sum_{i=0}^{p}\beta_i\right\}\Big/\left\{T_0\prod_{l=0}^{p}T_{r,l}\right\},$$

where $T_0 = \sum_{i=0}^{p}\exp\beta_i$ and $T_{r,l} = \sum^{(r,l)}\exp\beta_i$. The summation sign $\sum^{(r,l)}$ denotes that in the summation those terms corresponding to the first l elements of the rth permutation of 0 to p have been omitted. The conditional log likelihood of the observed results is given by

$$L(\boldsymbol{\beta}) = n\sum_{i=0}^{p}\beta_i - n\log T_0 - \sum_{r=1}^{(p+1)!} n_r \sum_{l=1}^{p}\log T_{r,l},$$

where n_r is the number of blocks in which the rth permutation occurs. It may be shown that

$$U_k \equiv \frac{\partial L(\mathbf{0})}{\partial \beta_k} = \sum_{l=1}^{p} \frac{m_{k,l}}{p-l+1} - n \sum_{l=1}^{p} \frac{1}{l+1},$$

where $m_{k,l}$ is the number of blocks in which the kth treatment ($k = 1, 2, \ldots, p$) has a rank of at most l. It may also be shown that

$$E\left\{\frac{\partial^2 L(\mathbf{0})}{\partial \beta_k^2}\right\} = -np\phi(p)$$

and

$$E\left\{\frac{\partial^2 L(\mathbf{0})}{\partial \beta_h \, \partial \beta_k}\right\} = n\phi(p),$$

where

$$\phi(p) = \frac{p+2}{(p+1)^2} - \frac{1}{p(p+1)} \sum_{t=1}^{p} \frac{1}{t},$$

so that the information matrix is given by

$$\mathbf{I} = n\phi(p) \begin{bmatrix} p & -1 & -1 & \ldots & -1 \\ -1 & p & -1 & \ldots & -1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -1 & -1 & -1 & \ldots & p \end{bmatrix}$$

with inverse

$$\mathbf{I}^{-1} = [n(p+1)\,\phi(p)]^{-1} \begin{bmatrix} 2 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 1 & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & \ldots & 2 \end{bmatrix}.$$

On the hypothesis that $\boldsymbol{\beta} = \mathbf{0}$, the test statistic

$$\mathbf{U}'\mathbf{I}^{-1}\mathbf{U} = 2[n(p+1)\,\phi(p)]^{-1} \sum_{h \leqslant k} U_h \, U_k$$

has, asymptotically, a χ^2 distribution with p degrees of freedom.

This statistic is quite different from that due to Friedman, which would usually be employed in this situation. As an example of its application Bradley (1968, Example 5.12.6, p. 127) gives data of the effect of four drugs on a person's visual acuity based on tests on five people. The rankings are as follows:

	Drug			
Subject	0	1	2	3
A	2	4	1	3
B	3	4	2	1
C	4	3	2	1
D	3	4	1	2
E	4	2	1	3

The table of the numbers $m_{k,l}$ together with the resulting values of U_k is:

k \ l	1	2	3	U_k
1	0	1	2	−35/12
2	3	5	5	37/12
3	2	3	5	21/12

This gives

$$\mathbf{U}'\mathbf{I}^{-1}\mathbf{U} = 1782/230 = 7{\cdot}75,$$

while the 5 per cent point of χ^2 with three degrees of freedom is 7·815. On the other hand Friedman's test for these data gave a value of 8·28 for a different approximate χ^2 variable with three degrees of freedom. These results are broadly in agreement.

It should be pointed out that the non-parametric analysis given here also provides in the statistics U_k some information about whether a treatment is "good" or "bad" relative to the standard. By inspection of those statistics treatment 1 is "bad", whereas 2 and 3 are "good". We can also attribute a standard error to the statistics U_k, given, asymptotically, by

$$\sqrt{[2/\{n(p+1)\,\phi(p)\}]}.$$

In the example this takes value 0·79. Because it is a fairly small experiment and because there is a high correlation between the U_k's, we need to adopt a cautious attitude in interpreting this standard error.

In principle this approach may be extended to deal with ties and/or with blocks (either incomplete or over-complete) of different sizes, although the algebra may not come out so neatly. By a suitable choice of "blocks" a non-parametric test may be derived for any situation, in which an analysis of variance test would be appropriate on continuous measurements. In particular a relatively simple test emerges for the k-sample situation (as an alternative to the usual Kruskal–Wallis test). For $k = 2$ this reduces of course to the test given in equations (32)–(36) of the present paper. This two-sample test was earlier described by Professor Cox in his 1964 paper as an example of the use of exponential scores. In fact all the tests developed by the method I have described can be expressed in terms of exponential scores, illustrating the point that the use of these scores arises from the Lehmann alternative rather than from the exponential distribution itself.

A rather more interesting and relatively simple non-parametric test that can be derived is for the equivalence of treatment effects in a balanced incomplete block experiment. Apart from its practical uses it is interesting because if there are only two treatments per block we are back again in a paired comparison situation, only this time paired comparisons of the Round Robin type. For this case Professor Cox's approach leads to the test given by David (1963, p. 38). Thus the methods of this paper applied to traditional non-parametric problems enable us to put under a single umbrella apparently unconnected situations.

As usual the statistical ideas that Professor Cox has discussed are of both theoretical interest and great practical importance. It gives me the greatest pleasure to propose the vote of thanks.

Mr Richard Peto (Oxford University): I have greatly enjoyed Professor Cox's paper. It seems to me to formulate and to solve the problem of the regression of prognosis on other factors perfectly, and it is very pretty.

In one detail I think that Professor Cox has not claimed the full credit that his method deserves. Suppose we have a single explanatory variable z and a single parameter β relating z to prognosis (i.e. to the distribution of failure time) and suppose that censoring is independent of z. In this situation, Professor Cox suggests in equation (18) the statistic $U(0)$ for testing $\beta = 0$. This test statistic is not merely *asymptotically* efficient, it is *locally most powerful* among all rank-invariant test procedures. This is exactly true for any particular finite sample size, and $U(0)$ is therefore the best conceivable rank-invariant test statistic for this problem.

In the case where z is a zero–one indicator variable, the test of $\beta = 0$ is the two-group rank test of Section 7, which is the logrank test and which has already been proved to be of maximal local power for detecting a difference between two groups of similarly censored observations. However, the discovery of a rank test of maximal local power for detecting

dependence of prognosis on a *continuous* variable is completely novel. We have used Professor Cox's regression methods in Oxford on real data and, despite appearances, they are computationally very quick and easy to handle, given careful programming.

I think only that his treatment of tied ranks is unsatisfactory. From the viewpoint of the analysis of clinical trials, it falls between two stools. His suggested likelihood function for tied ranks is not exactly the correct likelihood function if time is continuous and tied ranks merely represent slight grouping, although the exactly correct function is horribly complicated. However, if Cox's suggested likelihood function is seen as merely a very good approximation to the proper grouped-continuous-time likelihood function, then it can be shown that an equally good approximation, which is much simpler, exists (see below).

Now, it is not fair to complain that a paper which has been very full and interesting does not give all the techniques required for the analysis of clinical trials. However, it does seem to us at Oxford that a synthesis of Professor Cox's fully conditional regression and our fully permutational two-group significance testing is better than either separately.

In a clinical trial, patients are allocated at random to receive drug A or drug B and, as they enter the trial, various explanatory variables are recorded; white blood count, age and so on. Suppose we have a vector $\mathbf{z}$ of information on each patient, where z_1 is a zero–one indicator variable specifying group membership. Let $\boldsymbol{\beta}$ be the vector of coefficients relating to prognosis in exactly the manner Professor Cox has described. Professor Cox has suggested the following test for whether, after allowing for everything else, group membership affects prognosis. First, find $\hat{\boldsymbol{\beta}}_r$, the restricted ML value of $\boldsymbol{\beta}$ in which β_1, the group membership parameter, is constrained to be zero. Then examine what Professor Cox calls $U(0)$, which is the log-likelihood derivative at $\hat{\boldsymbol{\beta}}_r$ with respect to the group membership parameter β_1. Following Professor Cox, either the square root of the log likelihood increase when the restriction on $\hat{\boldsymbol{\beta}}$ is lifted or $U(0)$ is approximately normally distributed, and since z_1 is independent of the other components of $\mathbf{z}$ it does not matter which we examine to test whether treatment matters.

However, if this test is the heart of a clinical trial which has lasted several years, it is better for it to be exact than approximate. Having located the restricted likelihood maximum at $\hat{\boldsymbol{\beta}}_r$, we can in fact construct a *score* for each subject, expressing how well he has done given his initial white blood count, age and so on, such that the sum of the scores of the subjects in group A equals $U(0)$. The null distribution of $U(0)$ is therefore that of the sum of a random selection from the finite population of our derived scores, and exact significance tests are therefore possible.

Define the observed death count for subject j to be 1 or 0 according to whether the subject died or not, and define the expected death count for subject j to be an appropriate function of $\hat{\boldsymbol{\beta}}_r$,

$$\sum_{i\,|\,j\in R_{t(i)}} \exp(\hat{\boldsymbol{\beta}}_r . \mathbf{z}_j) \Big/ \sum_{k\in R_{t(i)}} \exp(\hat{\boldsymbol{\beta}}_r . \mathbf{z}_k),$$

which equals the risk of death on a man-years basis for subject j if the explanatory variables affect prognosis as $\hat{\boldsymbol{\beta}}_r$ (where for typographical reason $R_{(t_{(i)})}$ is printed $R_{t(i)}$).

The score for subject j is now the difference between his observed and his expected death-counts, and the sum of the scores for one particular treatment group equals plus or minus $U(0)$. The exact null distribution of $U(0)$ is therefore that of the sum of a random combination of these scores.

We have also found the calculation of observed and expected death-counts for individuals to be good for illustrating the dependence of prognosis on a particular factor. If the factor is divided into a few sub-groups, and the sums of the observed and the sums of the expected death-counts in those sub-groups are compared, then it is easier to understand physically the apparent nature of the dependence than if we just have a few regresssion coefficients and significance levels to look at.

Finally, I would like to return to the question of how Professor Cox deals with tied ranks when time is continuous and tied ranks mean only that slight grouping has occurred. If β is a vector of coefficients and $\mathbf{z}_j$ is the vector of explanatory variables for subject j, denote by e_j the quantity $\exp(\beta.\mathbf{z}_j)$. Also, I restrict attention to one event only (consisting of one death or several tied deaths), and abbreviate "the sum over the risk set of" to "the sum of". Now, at any particular time the death rate for subject j is proportional to e_j, so if one death only occurs the probability that it was subject j who died is $e_j/\sum e$. What likelihood should replace $e_j/\sum e$ if more than one death occurs? As Professor Cox remarks, any relatively *ad hoc* modification of his procedure will deal satisfactorily with this problem if the ties are few in number.

I will take the special case of two subjects, $j1$ and $j2$, dying at the same recorded time: generalization to several deaths is straightforward. If time is continuous, the probability that $j1$ and $j2$ are the two subjects who die is the sum of the probability that $j1$ dies first and $j2$ second plus the probability that $j2$ dies first and $j1$ second. Call this the *real probability*;

$$P_{\text{real}} = \frac{e_{j1}}{\sum e}\frac{e_{j2}}{(\sum e)-e_{j1}}+\frac{e_{j2}}{\sum e}\frac{e_{j1}}{(\sum e)-e_{j2}}.$$

Professor Cox's suggested probability appears in his equation (22); call this *Cox's probability*;

$$P_{\text{Cox}} = \frac{2e_{j1}\,e_{j2}}{(\sum e)^2-\sum e^2}.$$

I would like to suggest a third form that the probability might take, which I call the *rough probability*;

$$P_{\text{rough}} = e_{j1}\,e_{j2}\Big/\binom{N}{2}(\sum e/N)^2.$$

Physically, it is a matter of indifference which of the three forms we adopt. All are identically equal in the absence of tied ranks, and if there are tied ranks the differences between the three forms are two orders of magnitude less than the random variation which is being analysed. The rough probability is just as good an approximation to the real probability as Cox's probability, but all things being equal I suppose one would marginally prefer to use the real probability since no approximation to reality is involved. However, all things are not equal; the location, even given extremely efficient programming, of the maxima of likelihoods derived from the real probability or from Cox's probability is much more complex than the location of the maximum of the rough probability. For this reason, I believe that Professor Cox's model should perhaps be fitted in continuous time by maximizing the sum over all events of the logs of the rough probabilities. Susannah Howard has developed an algorithm which converges in powers of ten or better, and which is fast—the fit of five factors to 250 patients took less than a second per step on an old Atlas, and is, therefore, quite practicable.

Last week, I used these methods on some clinical trial data, and while I was going over the results someone asked me why I was looking so pleased. I said that it was because the method that was being used was so neat, and she asked me to explain it. She is not a mathematician nor a statistician, so I described the conditional argument and left out all the computational details. When I had finished, she said "I can't see why you think that's neat. It's just common sense." I second the vote of thanks to Professor Cox because he has opened up new territories to common sense.

The vote of thanks was put to the meeting and carried unanimously.

Professor D. J. BARTHOLOMEW (University of Kent): Professor Cox's methods have interesting potential applications to the analysis of labour wastage. The function $\lambda(t; \mathbf{z})$ then represents an individual's propensity to leave as a function of his length of service.

The form of this function has an obvious relevance to personnel policies and it has been the subject of a good deal of empirical work. Forbes (1971) reviewed the application of life table techniques to the non-parametric estimation of the survivor function from censored data.

It is well established that propensity to leave depends on many attributes of which sex, grade, level of skill, place of residence are among the most important. The methods given in this paper offer the prospect of a much more efficient estimation of these relationships than has hitherto been possible. The model of equation (9) is particularly appealing because of its simplicity and because of a certain plausibility which it has in the wastage application. The form of survivor functions is remarkably stable and this might suggest a common $\lambda_0(t)$ scaled up or down by a factor depending on the explanatory variables $\mathbf{z}$. Unfortunately there is a considerable body of empirical evidence to suggest that this is not the case. Survivor functions are often close to the lognormal implying that

$$\lambda(t) = \phi\left(\frac{\log t - \mu}{\sigma}\right) \Big/ \left\{1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right\},$$

where ϕ is the standard normal density and Φ its integral. Further, the parameter σ appears to reflect the type of job concerned (e.g. professional, skilled manual) whereas variation in the explanatory variables listed above exert their influence through μ. A suitable model might then be obtained by writing $\mu = \mathbf{z}'\boldsymbol{\beta}$ in $\lambda(t)$. The analysis could then be developed using parametric maximum likelihood methods but the simplicity of the author's methods would be lost. It would be interesting to know whether the methods of the paper are robust enough to give sensible answers when the lognormal model is appropriate. Put another way we might ask whether it is possible to construct $\mathbf{z}'\boldsymbol{\beta}$ in such a way that there is close agreement between the two models. Some of the z's would be the explanatory variables in which we are interested and others might be functions of t designed to improve the approximation.

The non-parametric estimation of survivor functions when $\lambda(t)$ is monotonic, referred to in Section 2, has been extended to increasing failure rate average (IFRA) distributions. A review of this general problem is to appear in Barlow *et al.* (1972).

Mr DAVID OAKES (Imperial College, London): I should like to remark briefly concerning the estimation of the distribution of failure time once an estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained. The method given in Section 8 of the paper treats $\lambda_0(t)$ as identically zero except at points where failures occur. However when dealing with data in continuous time it seems more natural to assume that $\lambda_0(t)$ is a slowly varying function of t. This leads to a simple maximum likelihood estimate of λ_k, the (assumed constant) value of $\lambda_0(t)$ between the failure times $t_{(k-1)}$ and $t_{(k)}$ ($t_{(0)} = 0$). We obtain

$$\hat{\lambda}_k = \left[\int_{t_{(k-1)}}^{t_{(k)}} \sum_{i=1}^{n} H(\tau_i - u) \exp\{\hat{\boldsymbol{\beta}}\mathbf{z}_i(u)\}\, du\right]^{-1},$$

where τ_i is the time to failure or censoring of the ith individual and $H(x)$ is the Heaviside unit function. In order to obtain a good indication of the behaviour of $\lambda_0(t)$ it will be necessary to apply some grouping or smoothing procedure to these estimates.

Professor D. V. LINDLEY (University College London): For simplicity, my remarks are confined to the two-sample problem in continuous time. Let sample 0 have m observations occurring at times $s_1, s_2, \ldots, s_m$ (either failures or censored), and let m' ($\leqslant m$) of them be failures. The corresponding data for sample 1 are n times $t_1, t_2, \ldots, t_n$ of which n' ($\leqslant n$) are failures. If $\mathscr{F}_i(t)$ are the survivor functions ($i = 0, 1$), $f_i(t) = -d\mathscr{F}_i(t)/dt$ the corresponding density functions and $\lambda_i(t)$ the hazard rates, so that $f_i(t) = \mathscr{F}_i(t)\,\lambda_i(t)$, each censored value contributes a term $\mathscr{F}(t)$, and each failure a term $\mathscr{F}(t)\,\lambda(t)$, to the

overall likelihood (as distinct from Cox's *marginal* likelihood). Hence the likelihood function is

$$\prod_{i=1}^{m} \mathscr{F}_0(s_i) \prod_{i \in F_0} \lambda_0(s_i) \prod_{j=1}^{n} \mathscr{F}_1(t_j) \prod_{j \in F_1} \lambda_1(t_j),$$

where F_i is the set of failures for sample i. If we write, with the author, $\lambda_1(t) = \psi\lambda_0(t)$, so that $\mathscr{F}_1(t) = \mathscr{F}_0(t)^{\psi}$, this becomes

$$\left\{\prod_{i=1}^{m} \mathscr{F}_0(s_i) \prod_{i \in F_0} \lambda_0(s_i) \prod_{j \in F_1} \lambda_0(t_j)\right\} \left\{\prod_{j=1}^{n} \mathscr{F}_0(t_j)^{\psi}\, \psi^{n'}\right\}.$$

Now $\lambda_0(t)$, and hence $\mathscr{F}_0(t)$, is unknown, so we should properly write $\lambda_0(t \mid \theta)$ and $\mathscr{F}_0(t \mid \theta)$ indicating a parametric dependence on θ, say. It is immediately apparent from the second set of braces that the obvious conditions for a marginal likelihood argument, namely that the likelihood factorizes into one part involving ψ, the parameter of interest, and another with θ, the nuisance parameter, does not obtain. So Cox's argument cannot be supported this way.

Suppose we take the case $\lambda_0(t) = \theta$, a constant. Then the likelihood is easily found to be

$$e^{-\theta(S+\psi T)}\, \theta^{m'+n'}\, \psi^{n'},$$

where $S = \sum_{i=1}^{m} s_i$ and $T = \sum_{j=1}^{n} t_j$. If the prior is proportional to $\theta^{-1}\psi^{-1}$, we easily obtain the posterior for ψ to be proportional to

$$\psi^{n'-1}/(S+\psi T)^{m'+n'} \qquad (*)$$

so that $\psi\bar{t}/\bar{s}$ is F on $(2n', 2m')$ d.f.: here $s = S/m'$ and $t = T/n'$. (Notice the division by m', n'; not m, n.)

However the assumption of constant hazard is not necessarily appropriate, and is clearly avoided in the marginal likelihood approach. But for any $\lambda_0(t)$ there is a transformation of the time axis so that it is constant and again (*) will obtain but with S and T now the sums on the new time scale. Hence we can explore a range of prior estimates for $\lambda_0(t)$ and see how the results are affected.

It is worth contrasting the marginal likelihood with the integrated (with respect to θ) likelihood, equal to (*) times ψ. The former is a product of terms like $\psi/(a_i + b_i\psi)$ or $(a_i + b_i\psi)^{-1}$ where a_i and b_i refer to the numbers at risk. The numerators are at most different by ψ but the denominators are quite different since the times appear in (*) but not in the marginal likelihood. Special cases are worth exploring. Suppose sample 0 has one censored value at 2, and sample 1 has a failure at 1. Then the marginal likelihood is $\psi/(1+\psi)$ referring to the single risk set at $t = 1$. The integrated likelihood is $\psi/(2+\psi)$. With a change of time scale the most that the latter could be is $\psi/(1+\psi)$, and this when $t = 2$ is identified with $t = 1$. The marginal likelihood is therefore very extreme, especially in its failure to depend on the time of censoring or failure in sample 0 whenever this exceeds 1.

Mr P. W. GLASSBOROW (British Rail): I want to make a brief remark. In Section 8 Professor Cox analyses two causes of failure and whether the causes of failure are independent. In real life they often are not independent and this brings us back to the beginning of the paper. It is unfortunate that Professor Cox uses the term "censored"; I do not know whether this has been used elsewhere instead of the traditional term "withdrawal". If you use "withdrawal" you realize it is just a type of failure, and withdrawal and failure are often not independent.

The following contributions were received in writing after the meeting.

Professor D. E. BARTON (Queen Mary College and Institute of Computer Science, University of London): My feeling is that Professor Cox understates the importance of

Kaplan and Meier's result that the product-limit estimate is the maximum likelihood one and, conversely, is too kind to those who find the analytic problems in specifying the family of all possible distributions. As discussed in Barton (1968) there are several possible alternative forms of estimator, and it is not immediately clear that the maximum likelihood estimator makes best use of the information available. Moreover there are effectively an infinite number of nuisance parameters being eliminated (that is a nuisance function: the unknown censoring rule). In the paper cited I show that the method of maximum likelihood gives more efficient estimation than the alternatives and a heuristic argument suggesting that it is *efficient*. This efficiency does seem to be a property which gives Kaplan and Meier's result some importance.

Miss SUSANNAH HOWARD (Department of Biomathematics, Oxford University): Since Professor Cox has proposed such a satisfying method for the analysis of censored failure times, it seems worth while indicating how easily the computation involved can be performed.

By replacing the explanatory variables $\mathbf{z}_j$ for each individual j by $\mathbf{z}_j - \bar{\mathbf{z}}$, where $\bar{\mathbf{z}}$ is the mean of $\mathbf{z}$ over all those individuals who are observed to fail, the term $\sum \mathbf{s}_{(i)}\boldsymbol{\beta}$ in the full conditional log likelihood (following equation (22)) vanishes identically, giving

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

in either discrete or continuous time. The notation here is as in Section 6 of the paper, but with $\bar{\mathbf{z}}$ now equalling $\mathbf{0}$. If there are ties, L and its first and second derivatives can be computed by exploiting their "symmetric function" properties in the following way.

Let e_j be the exponential weight $\exp\{\mathbf{z}_j\boldsymbol{\beta}\}$ for the jth individual, and, for $1 \leqslant \xi, \eta \leqslant p$, define

$$x_{\xi j} = z_{\xi j}\, e_j, \quad y_{\xi\eta j} = z_{\xi j}\, z_{\eta j}\, e_j.$$

For any risk set $\mathscr{R}$ and any integer m, define $a(\mathscr{R}; m)$, $b(\mathscr{R}; \xi; m)$, $c(\mathscr{R}; \xi, \eta; m)$ and $d(\mathscr{R}; \xi, \eta; m)$, for $1 \leqslant \xi, \eta \leqslant p$, by recursion on $\mathscr{R}$:

(i) If $\mathscr{R} = \emptyset$,

$$a(\mathscr{R}; m) = \delta_{0,m},$$

and

$$b(\mathscr{R}; \xi; m) = c(\mathscr{R}; \xi, \eta; m) = d(\mathscr{R}; \xi, \eta; m) = 0 \quad \text{for all } m.$$

(ii) If $\mathscr{R}^+ = \mathscr{R} \cup \{j\}$, with $j \notin \mathscr{R}$,

$$a(\mathscr{R}^+; m) = a(\mathscr{R}; m) + e_j\, a(\mathscr{R}; m-1),$$

$$b(\mathscr{R}^+; \xi; m) = b(\mathscr{R}; \xi; m) + e_j\, b(\mathscr{R}; \xi; m-1) + x_{\xi j}\, a(\mathscr{R}; m),$$

$$c(\mathscr{R}^+; \xi, \eta; m) = c(\mathscr{R}; \xi, \eta; m) + e_j\, c(\mathscr{R}; \xi, \eta; m-1) + y_{\xi\eta j}\, a(\mathscr{R}; m),$$

$$d(\mathscr{R}^+; \xi, \eta; m) = d(\mathscr{R}; \xi, \eta; m) + e_j\, d(\mathscr{R}; \xi, \eta; m-1) + x_{\xi j}\, b(\mathscr{R}; \eta; m) + x_{\eta j}\, b(\mathscr{R}; \xi; m)$$

for all m.

Then

$$L = -\sum_{i=1}^{k} \log a_i \quad \text{where } a_i = a(\mathscr{R}(t_{(i)}); m_{(i)}),$$

$$\frac{\partial L}{\partial \beta_\xi} = -\sum_{i=1}^{k} b_{\xi i} \quad \text{where } b_{\xi i} = b(\mathscr{R}(t_{(i)}); \xi; m_{(i)} - 1)/a_i$$

and

$$\frac{-\partial^2 L}{\partial \beta_\xi\, \partial \beta_\eta} = \sum_{i=1}^{k} \{c_{\xi\eta i} - b_{\xi i}\, b_{\eta i}\},$$

where

$$c_{\xi\eta i} = \{c(\mathscr{R}(t_{(i)});\ \xi, \eta;\ m_{(i)}-1) + d(\mathscr{R}(t_{(i)});\ \xi, \eta;\ m_{(i)}-2)\}/a_i.$$

Now if we consider the times of censoring or observed failure in reverse chronological order, the risk set $\mathscr{R}$ increases steadily. So after first defining arrays $A(m)$, $B(\xi, m)$, $C(\xi, \eta, m)$ and $D(\xi, \eta, m)$ according to (i), with ξ, η and m within the following bounds

$$\left.\begin{array}{ll} A(m) & 0 \leqslant m \leqslant m_\infty \\ B(\xi, m) & 1 \leqslant \xi \leqslant p,\ 0 \leqslant m \leqslant m_\infty - 1 \\ C(\xi, \eta, m) & 1 \leqslant \xi \leqslant \eta \leqslant p,\ 0 \leqslant m \leqslant m_\infty - 1 \\ D(\xi, \eta, m) & 1 \leqslant \xi \leqslant \eta \leqslant p,\ 0 \leqslant m \leqslant m_\infty - 2 \end{array}\right\} \text{where } m_\infty = \max\{m_{(i)} \mid 1 \leqslant i \leqslant k\},$$

then as each new individual joins the risk set the corresponding values in the arrays can be computed according to (ii). Thus at each failure time $t_{(i)}$ all the terms needed for computing L and its derivatives are already known. Moreover, at any time t, in the bounds given above, m_∞ may be replaced by $m_t = \max\{m_{(i)} \mid t_{(i)} \leqslant t\}$.

This simple procedure can be programmed in a way which allows for flexibility, so that one can choose whether or not to use approximations for the second derivatives, or even for L itself. If there are not too many data (say, up to 200 individuals with not more than 5 parameters to be fitted), maximization of the full conditional log likelihood is feasible without resorting to approximations and, in situations where time is genuinely discrete, as one might find in certain types of life-testing, it is better to fit the logistic model exactly. However, in analysing a large clinical trial with "ties" due to slight grouping, approximations such as the "rough" probability which Richard Peto has suggested would still seem preferable.

Professor B. BENJAMIN (Civil Service College): It is not quite true that actuaries are only concerned with situations in which sampling errors are insignificant. Many of them are involved in follow-up studies of special groups (e.g. those with impairments) or with non-life investigations which are analogous to reliability trials. The actuary is moreover not only interested in the probability of surviving t years, or the expectation of life, or the expected number of "failures" in a specified period. He is interested in the *shape* of the life table. He is a collector of shapes and part of his special skill lies in his experience of and recognition of typical shapes. My approach to the data of Table 1 is as follows. (1) Turn the table upside down and group in 5-week periods to reduce irregularities. Assuming that the failures are at the nearest integral interval and that the censored "lives" survived to the beginning of the interval in which they were censored, calculate the average exposed and thence the average death-rates m_t in each interval [note that we do not wholly discard the censored "lives"]. (2) Plot these and draw a smooth curve through the points (see Fig. I) thus inferring and removing sampling fluctuations (there are tests for improving the efficiency of this inference—see Benjamin and Haycocks, 1971). The shape of m_t is reminiscent of many curves with a basic exponential progression and an additional component of early "mortality" probably like some population life tables where m_x, x being age, is a combination of a Gompertz ($m_x = Bc^x$) and a Normal curve in early ages. It is also very evident without calculating errors that the two experiences are different. I have not seen the author's diagram. (3) Read off m_t for each week and calculate first p_t and then $[p_0.p_1.p_2.\ \ldots\ .p_{t-1}]$ the probability of surviving t intervals. (4) Calculate the variance of this probability and if necessary make a formal test of the difference between the two experiences. In this case, as the author agrees, the difference is overwhelmingly significant. There are probably weaknesses and strengths in this procedure. The author would be doing the actuaries a great service if he would turn to a practical review of these weaknesses and to an assessment of their importance in practical situations (like Table 1). Actuaries are willing and able to follow the mathematics especially when so lucidly expressed as in this paper but they need to be convinced that it is important to decision-taking.

For what it is worth my estimates for the proportion surviving 5 and 10 weeks respectively are (variance in brackets)

Sample I 0·649 (0·0155) 0·358 (0·0569),

Sample O 0·923 (0·0040) 0·753 (0·0145).

May I also stress, as elsewhere (Benjamin, 1972), that no actuary would recommend *action* on any experiment for which significance could be demonstrated only after great mathematical strain. Most important changes stick out like a sore thumb.

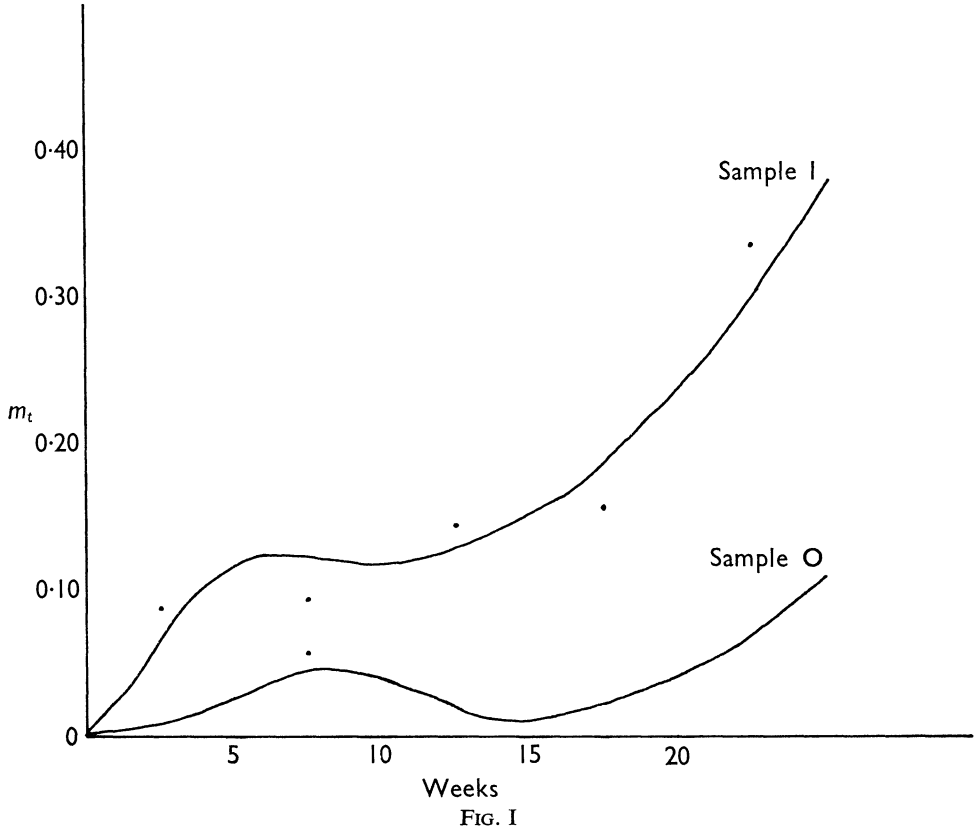

FIG. I

Dr JOHN J. GART (National Cancer Institute): In 1958 Professor Cox presented an elegant and unified approach to the analysis of binary data and now he gives a treatment of life tables of equal elegance and usefulness. In Section 7 he points out the formal identity of (27) to the test for partial association in combining 2×2 contingency tables. It follows almost as directly that the χ^2 test statistic for the comparison of $p+1$ independent survival curves derived from (18) is formally identical to the Birch–Armitage statistic for partial association in $2 \times k \times (p+1)$ contingency tables (Birch, 1965; Armitage, 1966). In the two-sample problem, it appears that valid, asymptotic methods for the point and interval estimation of e^{β} are formally identical to those of the common odds ratio in combining 2×2 contingency tables (e.g. Gart, 1970). It will prove interesting to pursue further the possible parallels between life tables and contingency tables. Can the formally identical tests for interaction in higher dimensional contingency tables be used to test the

plausibility of the proportional hazard rate model? Will the proportional hazard rate model methods prove as robust as the logistic model methods for contingency table analyses? Once again Professor Cox has provided a simple, coherent framework within which such questions can be resolved.

Drs L. D. MESHALKIN and A. R. KAGAN (World Health Organization): We congratulate the author on an extremely stimulating paper, which has relevance to epidemiological studies of prediction of high risk and identification of causes, as well as to clinical trials. We make below two points, illustrated by an example of how the power of a particular factor (raised blood pressure) to predict subsequent disease (death from cardiovascular disease) varies with the interval between its measurement and the onset of disease. We believe that this demonstrates further the ideals expressed by Cox.

1. Use of a more complicated function $h(z, \beta)$. Predictors of those at high risk to develop ischaemic heart disease have been identified by relating initial measurements made on groups of subjects to their subsequent disease experience. But the predictive power of some factors changes with the passage of time. It is important to know the way in which this change takes place for a proper understanding of the disease process and its control and also for more adequate study design.

An adaptation of Professor Cox's approach enables us to measure this even when the study includes subjects of different age, who remain in the study for varying periods of time and the number of subsequent disease events is small (e.g. 684 males were followed for not more than 10 years, aged 30–62 years at entry, with 66 cases of cardiovascular death).

Our illustration (Fig. II in this Discussion) shows how the predictive power of the value of the systolic blood pressure decreases. Two analytical expressions were used for the function, $h(z, \beta)$:

$$h_1 = (\beta_0 + \beta_1 z)(1 - \beta_2)^\tau,$$

$$h_2 = (\beta_0 + \beta_1 z)/(1 + \beta_2 \tau),$$

where τ is a time from the initial measurement and z the value of a systolic blood pressure. Fig. II shows that the choice of analytical expression has not influenced the result much.

2. A knowledge of $\lambda_0(t)$. For a number of chronic diseases, $\lambda_0(t)$ can be well approximated by the function,

$$\lambda_0(t) = \exp\{d_0 + d_1 t\}$$

as used, for example in de Haas (1964).

In the above example, use of this form of function $\lambda_0(t)$ reduces asymptotic variances of estimates by 10–20 per cent.

Computer programs for the above analyses can be obtained from the Numerical Analysis Unit of the Division of Research in Epidemiology and Communications Science, of the World Health Organization, Geneva, Switzerland.

Professor M. ZELEN (State University of New York at Buffalo): My congratulations to Professor Cox on presenting a very stimulating and pioneering paper. He has raised several points in his paper which I am certain will be the subject of much future investigation. I wish to confine my remarks to the analogy between the model discussed by Professor Cox and contingency tables. To simplify matters only the two-sample problem will be discussed and no censoring will be assumed present.

Suppose we have $(k+1)$ intervals $(z_{\alpha-1}, z_\alpha]$ $(\alpha = 1, 2, \ldots, k)$, (z_k, ∞) where $z_0 = 0$. Also let there be two populations having the conditional probabilities $p_{i\alpha} = \mathscr{F}_i(z_\alpha)/\mathscr{F}_i(z_{\alpha-1})$ for $i = 1, 2$. (Choose z_k so that there are no failures past z_k.) Then, if the event of surviving or not surviving an interval is only considered for analysis, the comparison of the two populations is formally the same (as Professor Cox has noted) as comparing several 2×2

contingency tables. The test statistic depends on the alternative hypothesis whether the odds ratio $\psi_\alpha = q_{1\alpha}p_{2\alpha}/q_{2\alpha}p_{1\alpha}$ for the αth table are the same or possibly different. If $\psi_\alpha = \psi$ for all α, then the appropriate test statistic is the one discussed by Cochran (1954) and Mantel and Haenzel (1959). Alternatively, if ψ_α are not all equal the test statistic would be

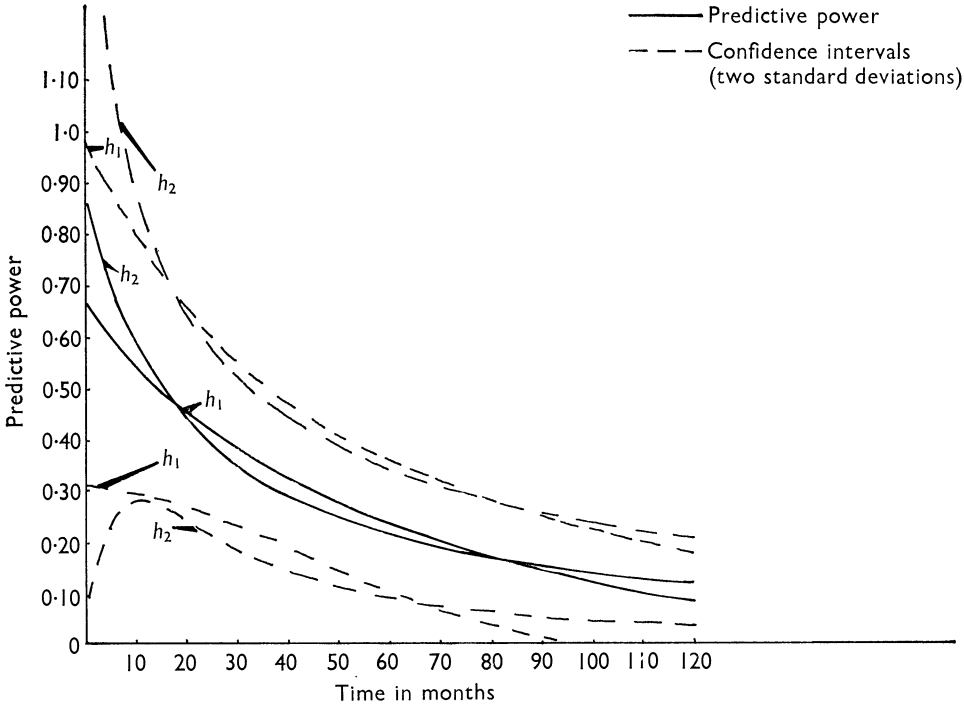

Fig. II. Predictive power of initial systolic blood pressure as a function of time from initial measurement.

Predictive power is measured by:

$$\log_{10} h(x_1)/h(x_2),$$

where $h = h(x)$ is a factor which shows how many times the risk of an individual with a measurement value of x is more than the risk for the average individual of his age, and $x_1(x_2)$ is the value of measurement such that one-quarter of the whole population of his age has bigger (lower) values of x.

different, cf. Zelen (1971). For example, if the two populations have exponential distributions ($\mathscr{F}_i(t) = \exp - \lambda_i t$), we have

$$\psi_\alpha = [\{1 - \exp(-\lambda_1 \Delta_\alpha)\}/(1 - \exp(-\lambda_2 \Delta_\alpha)\}] \exp - (\lambda_2 - \lambda_1)\Delta_\alpha,$$

where $\Delta_\alpha = z_\alpha - z_{\alpha-1}$. Thus the ψ_α will not be the same (provided $\lambda_1 \neq \lambda_2$) unless the intervals are chosen to be of equal length. In general for arbitrary survival distributions where $\mathscr{F}_2(t) = [\mathscr{F}_1(t)]^p$, the same result will hold in that the $\{\psi_\alpha\}$ will be different. The same remarks hold if the intervals are chosen to coincide with the observed failure times. Thus the asymptotic test procedure will not in general lead to equation (27) of Professor Cox's paper.

Professor R. E. BARLOW (University of California at Berkeley): Professor Cox has proposed some apparently very useful procedures for analysing life test data. In a recent paper with Doksum (1972), we found the *cumulative total time on test statistic* to be very useful in the single sample goodness-of-fit problem for exponentiality. This statistic is

$$\sum_{i=1}^{k} \int_0^{X_{i:n}} [1 - F_n(u)]\, du \Big/ \int_0^{X_{n:n}} [1 - F_n(u)]\, du,$$

where F_n is the empirical distribution and $X_{1:n} \leqslant X_{2:n} \leqslant \ldots \leqslant X_{k:n}$ are the first k order statistics from F. The process

$$\int_0^{F_n^{-1}(t)} [1 - F_n(u)]\, du$$

on [0, 1] also played a key role in Barlow and van Zwet (1970) where we investigated estimates for the failure rate assumed monotone. These statistics thus seem useful in life test models besides those based on the exponential distribution. Perhaps since the present paper is more concerned with supplementary information, total time on test statistics does not play such a central role. However, I would like to see a formulation of these problems in which the total time on test statistics might be used to advantage.

Reference should perhaps be made to the relevant paper by Harris *et al.* (1950) in connection with step function failure rate estimators.

Doksum (1967) also uses tests based on (32) for non-parametric two-sample life test problems. He shows that the Savage statistic (32) maximizes the minimum power over IFRA (for increasing failure rate average) distribution, F, asymptotically, for the problem H_0: $\Delta \leqslant 1$ versus H_1: $\Delta > 1$ where the first sample is from $F(.)$ and the second sample is from $F(./\Delta)$.

Recently, some very elegant properties of shock model processes have been discovered by Esary *et al.* (1972). Perhaps, these are now ripe for statistical analysis.

Drs JACK KALBFLEISCH and R. L. PRENTICE† (State University of New York at Buffalo): We would like to raise some questions concerning the conditional likelihood in Section 5 of this paper. Let us suppose a continuous hazard without censored observations. Expression (12) appears to be the conditional probability that individual i fails at $t_{(i)}$, given that a failure occurs at $t_{(i)}$ and given the risk at $R(t_{(i)})$. Thus if individuals 1, 2, 3 have associated covariate values z_1, z_2, z_3 and are observed to fail at t_1, t_2, t_3, with $t_1 < t_2 < t_3$, then expression (12) yields

(i) P (1 fails at t_1 | one failure at t_1 and $R(t_1) = \{1, 2, 3\}$)
$= \exp\{z_1 \beta\} / \Sigma_1^3 \exp\{z_i \beta\}$;

(ii) P (2 fails at t_2 | one failure at t_2 and $R(t_2) = \{2, 3\}$)
$= \exp\{z_2 \beta\} / \Sigma_2^3 \exp\{z_i \beta\}$;

(iii) P (3 fails at t_3 | one failure at t_3 and $R(t_3) = \{3\}) = 1$.

Our questions concern the combination of such statements to form the expression (13). If (13) is the logarithm of a conditional likelihood, then the product of (i), (ii) and (iii) should permit an interpretation as a conditional probability statement. The introduction of Section 5 appears to suggest that the distribution to be calculated is to be conditional on the observed order statistic. However, the conditional portion of (i), for instance, is the event that a failure occurs at t_1 and two failures occur after t_1 (as opposed to the event that failures occur at t_1, t_2, t_3). Thus the likelihood corresponding to (13) differs from that arising from the permutation distribution calculated conditionally on the observed failure times. The permutation distribution generally involves $\lambda_0(t)$ ($\beta \neq 0$).

† On leave from the University of Waterloo, Canada.

On casual reading, it appears that (13) is formed by regarding the selection of an individual from the risk set at each observed failure time as an independent experiment. The Cartesian product of the conditional probability spaces corresponding to each such experiment would then give a probability yielding (13) as the log likelihood of β. This procedure, however, defines a reference set which attaches positive probability to events in which the same individual fails several times. We would appreciate it if Professor Cox would discuss the reference set and the conditional probability statements from which (13) arises.

Considering again the continuous uncensored case, it is of interest to note that the model (9) is invariant under the group of differentiable, monotone, strictly increasing transformations on survival time. This invariance permits the calculation of a marginal likelihood (Kalbfleisch and Sprott, 1970, or Fraser, 1968) for β. The marginal likelihood, the logarithm of which is given by (13), arises from the marginal distribution of the ranks. The continuous censored case can also be handled from the viewpoint of marginal likelihood by imposing approximations similar to those in Section 5. Again the resulting expression is (13). If multiplicities are allowed in the continuous case, the resulting marginal likelihood differs from (22) and is written

(iv) $$\sum_{i=1}^{k} s_{(i)}\,\beta - \sum_{i=1}^{k} m_{(i)} \log \sum_{j \in R(t_{(i)})} \exp\{z_j \beta\}.$$

Expression (iv) seems appealing in certain special instances considered. For example, if $n = 2$ and $t_1 = t_2$ is observed with corresponding covariate values z_1 and z_2, then (iv) has a unique maximum at $\beta = 0$ unless $z_1 = z_2$. Expression (22), however, reduces identically to zero in this case, indicating that no one value of β is to be preferred to any other. But, if z_1 and z_2 differ widely it seems clear that $\beta = 0$ is to be favoured (provided the intervals for measuring survival time are not unduly large).

In order to keep these comments relatively brief, the calculations involved in obtaining these marginal likelihoods have been deferred to a note now being prepared for publication.

A final question involves the specification of the continuous model (9). Professor Cox suggests that a function of survival time itself may be used as a covariate in the hazard function. Since no assumption is made about $\lambda_0(t)$, the hazard

$$\lambda(t, z) = \lambda_0(t) \exp\{\beta_1 z\}$$

may be re-written as

$$\lambda(t, z) = \lambda_1(t) \exp\{\beta_2 t + \beta_1 z\}$$

without additional assumption. Corresponding to these two specifications, different conditional likelihoods (13) could be formed, which would generally give rise to different estimates of β. We note that the above-mentioned marginal likelihoods do not permit the inclusion of such time dependent covariates, and we would appreciate a discussion of when such covariates should be included.

Professor NORMAN BRESLOW (University of Washington): Like some of the other discussants I too was puzzled by the conditional likelihood of Section 2. I would like to suggest an alternative approach to the estimation of $\boldsymbol{\beta}$ and λ_0 which leads to equation (14) and also to a simpler estimate of the underlying survival distribution than is provided by equations (37) and (38). This approach is motivated in part by the discussion of Kalbfleisch and Prentice. However it differs from both their arguments and those of Cox in that simultaneous estimation of $\boldsymbol{\beta}$ and λ_0 is achieved through consideration of a joint likelihood function involving both sets of parameters.

One of the methods of deriving the Kaplan–Meier estimate in a maximum likelihood (ML) framework is to restrict attention to distributions having a hazard function which is constant between the distinct observed uncensored failure times, i.e.

$$\lambda_0(t) = \lambda_i \quad \text{for } t_{(i-1)} < t \leqslant t_{(i)}, \quad i = 1, \ldots, k.$$

This is also the starting point from which Granander (1956) derives ML estimates in the class of distributions with monotone hazard functions. Writing down the joint likelihood for Cox's model with λ_0 as defined above, and adopting Kalbfleisch and Prentice's convention of considering all censored observations as censored at the preceding uncensored failure time, it turns out that the values of β and λ_i which simultaneously maximize the likelihood are given by setting Cox's equation (14) to 0 to find $\hat{\beta}$ and by

$$\hat{\lambda}_i = m_{(i)}/(t_{(i)} - t_{(i-1)}) \sum_{l \in R(t_{(i)})} \exp(z_l \hat{\beta}).$$

Hence the estimate of the cumulative hazard

$$\Lambda(t) = -\log \mathscr{F}(t) = \int_0^t \lambda(u)\, du$$

evaluated at $t_{(i)}$ is

$$\hat{\Lambda}(t_{(i)}) = \sum_{j=1}^{i} m_{(j)} \Big/ \sum_{l \in R(t_{(j)})} \exp(\mathbf{z}_l \hat{\beta}).$$

With $\hat{\beta} = 0$ this is the form of the Kaplan–Meier estimate considered by Nelson (1969). To achieve an exact analogue of the Kaplan–Meier estimate, one may take

$$\hat{\mathscr{F}}(t) = \prod_{t_{(i)} < t} (1 - \hat{\pi}_i)$$

where

$$\hat{\pi}_i = m_{(i)} \Big/ \sum_{l \in R(t_{(i)})} \exp(\mathbf{z}_l \hat{\beta}).$$

This expression for the $\hat{\pi}_i$ can also be obtained as a first-order approximation to the estimate suggested by Cox and, as noted by them, as an approximation to the estimate derived from the distinct discrete time model of Kalbfleisch and Prentice.

I have recently applied Cox's regression model to the covariance analysis of survival data arising from a clinical trial involving 268 patients on 5 regimens. When the estimate of the underlying survival distribution suggested above was compared to the more complicated estimate of Cox, the two were found to agree to within 0·001 at each time point. Even more surprising was the fact that neither departed greatly from the unadjusted Kaplan–Meier estimate, obtained by setting $\hat{\beta} = 0$ in the expression for $\hat{\pi}_i$ above. This was true in spite of the fact that the covariate had a marked effect on survival.

The AUTHOR replied briefly at the meeting and subsequently more fully in writing as follows.

I am very grateful to all the contributors for their constructive and helpful comments. Many points have been made and it is not feasible to comment on them all.

Professor Downton has discussed a number of interesting non-parametric procedures which have good properties when the data are derived from underlying exponential variates. One question here concerns whether it is practicable to test from data whether such tests are more appropriate than, say, those based on underlying normal variates.

Mr Peto has made a number of very cogent points. The fact that "exact" tests can be based on the permutation distribution, while it does require the extra assumption that consoring operates equally on all groups, is important. Also his suggestion of a simpler approximate likelihood for the grouped case is ingenious and should certainly be noted by anyone proposing to use these methods, as should Miss Howard's valuable contribution on computational methods.

Professors Lindley, Zelen, Breslow and Kalbfleisch and Prentice all raise questions about the likelihood (12). The paper is unduly cryptic over this and I agree that further work may be needed to clarify exactly what is being done. The essence of the argument seems to me to be as follows.

(a) If $\lambda_0(t)$ is specified parametrically, the ordinary likelihood is used consisting, when the explanatory variables are independent of time, of a product of density functions from the individuals who fail and survivor functions from the individuals who are censored. This can be regarded as an integral to which all elements of time at risk contribute.

(b) If $\lambda_0(t)$ is arbitrary, (a) is not helpful. (Professor Lindley's remark about a transformation of the time scale is, I think, useful only when $\lambda_0(t)$ is known.) We therefore consider the likelihood for a description of part of the data, namely the specification of those individuals who fail considering hypothetical repetitions in which the times of failure are fixed. The probabilities in this new random system are deduced from those in the original fuller specification. Each probability is conditional on what happened at the previous time-points and on any intervening censoring. Factors associated with non-occurrences in intervening time-intervals are, however, not included. This is in the spirit of Bartlett (1937).

(c) This raises a number of issues.

(i) It is assumed without proof in the paper that the usual asymptotic procedures and properties associated with maximum likelihood estimates and tests hold.

(ii) Is it possible and worth while to try to recover information which for any specific $\lambda_0(t)$ is contained in the gaps between failures?

(iii) What is the loss of information about the regression coefficients involved in using the procedures of the paper when some parametric representation of $\lambda_0(t)$ is in fact appropriate? This clearly depends on the magnitude of the regression effects present.

Both Professors Lindley and Zelen work with formulations in which an exponential assumption allows use of information arising from gaps. Their results therefore differ from the results of the paper which, at least when the expanatory variables are independent of time, are invariant under monotonic transformations of the time scale, a property emphasized by Mr Peto; see especially Peto and Peto (1972). Incidentally a non-Bayesian version of Professor Lindley's main result is used at the end of Section 10 in comparing alternative analyses.

Professor Breslow's interesting derivation is not, I feel, essentially different from what I have done. He attaches a separate unknown parameter to every gap. This is an oblique way of saying that the gaps contribute no information about β. His likelihood function has a very large number of unknown parameters and this is well known to be dangerous.

In discrete time the position is in some ways more complicated. The logistic model used in (21) is possibly sensible for a process "really" taking place in discrete time, but is only a first-order approximation when the data are obtained by grouping a process in continuous time to which (9) applies. Putting the same point another way, if we had large amounts of data from the same system in two sets with greatly different grouping intervals, slightly different estimates would be obtained for the regression coefficients. This is unlikely to be a serious practical point and from this point of view there being an approximation anyway, use of Mr Peto's simpler function seems entirely sensible.

Mr Oakes's suggestion appears superior to that of Section 8 of the paper.

Professor Bartholomew has raised some interesting questions, which serve in particular to emphasize that a simple model in terms of hazards may not be the best way to proceed. Dr Meshalkin and Dr Kagan's contribution is very welcome as illustrating both a more complicated form of dependence on the explanatory variables and the use of a parametric assumption for $\lambda_0(t)$.

Mr Glassborow stresses an important assumption about censoring. As to terminology, I think I have followed that usual in statistical papers although this may well not be ideal. It is worth emphasizing that the discussion of Section 9 is concerned with the possibly rather unusual situation where there are two or more distinct kinds of failure time, all of which may be observed, and not with the situation where only one kind of failure time can be observed on any one individual.

Professor Benjamin's analysis is not all that different from the one of the paper especially in the light of Fig. 1 (which unfortunately was not available at the meeting). His approach is in some ways simpler, and therefore better, than that of the paper. On the other hand, the regression approach deals more readily with complex problems involving many explanatory variables. Also in simpler problems, provided that the relation between the different hazards is fairly direct, the comparison between them is made concisely in terms of parameters with a quite immediate physical meaning. Of course I agree that in taking action one wants the statistical uncertainty in the narrow sense to be small, although there surely are situations where this is not achievable.

I agree with Professor Barton that the difficulty in specifying the space of distributions involved in the maximum likelihood property of the product-limit method is not to be taken very seriously. On the other hand the property is analogous to that for a multinomial distribution with a very large number of cells and typical observed occupancies all very small, and the usual justifications for maximum likelihood are then fairly irrelevant.

Dr Gart has raised the important possibility that a wide variety of contingency table techniques can be adapted. Professor Barlow mentions a number of very interesting recent investigations. It seems quite possible that time on test could be adapted to problems of this paper by working with an estimated operational time variable after preliminary estimation of the regression coefficients.

Professors Kalbfleisch and Prentice have asked for clarification of the role of time-dependent explanatory variables. These must be either fixed functions for each individual or, if random, we argue conditionally on their realized values. If we were to take the same fixed function for each individual, e.g. t itself, the contribution would disappear from (12), the function having been absorbed into $\lambda_0(t)$. In the example we have an explanatory variable that is t for some individuals and zero for others.

Finally I would like to stress that while the model (9) seems to provide a flexible and simple way of representing a wide range of situations it is only one such way and the possibility of other physically sounder or more economical models should not be overlooked. Further, given the model (9), the method of analysis given main emphasis here is only one way of proceding and the possibility of a parametric representation of $\lambda_0(t)$ will often be worth consideration.

References in the Discussion

Armitage, P. (1959). The comparison of survival curves. *J. R. Statist. Soc.* A, **122**, 279–300.

—— (1966). The chi-squared test for heterogeneity of proportions after adjustment for stratification. *J. R. Statist. Soc.* B, **28**, 150–163; Addendum: 1967, **29**, 197.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions.* Chichester: Wiley.

Barlow, R. E. and Doksum, K. (1972). Isotonic tests for convex orderings. In *Proc. 6th Berkeley Symp. on Math. Statist. Prob.*, pp. 293–323. Berkeley: University of California Press.

Barlow, R. E. and Van Zwet, W. (1970). Asymptotic properties of isotonic estimators for the generalised failure rate function. In *Proc. 1st Int. Symp. on Non-parametric Techniques in Statistical Inference*, pp. 159–174. Cambridge: University Press.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc.* A, **160**, 268–282.

Barton, D. E. (1968). The solution of stochastic integral relations for strongly-consistent estimators of an unknown distribution function from a sample subject to variable censoring and truncation. *Trab. Estadist.*, **19**, 51–73.

Benjamin, B. (1972). Stochastic aspects of life tables. *I.M.A. Bull.*, **8**, 12–16.

Benjamin, B. and Haycocks, H. W. (1971). *The Analysis of Mortality and other Actuarial Statistics.* Cambridge: University Press.

Birch, M. W. (1965). The detection of partial association, II. The general case. *J. R. Statist. Soc.* B, **27**, 111–124.

Bradley, J. V. (1968). *Distribution-free Statistical Tests.* Englewood Cliffs, N.J.: Prentice Hall.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, **10**, 417–451.

Cox, D. R. (1958). The regression analysis of binary sequences (with Discussion). *J. R. Statist. Soc.* B, **20**, 215–242.
David, H. A. (1963). *The Method of Paired Comparisons.* London: Griffin.
De Hass, J. H. (1964). *Changing Mortality Patterns and Cardiovascular Diseases.* N. V. Haarlem: De Erven F. Bohn.
Doksum, K. (1967). Asymptotically optimal statistics in some models with increasing failure rate averages. *Ann. Math. Statist.*, **38**, 1731–1739.
Esary, J. D., Marshall, A. W. and Proschan, F. (1972). Shock models. *Ann. Math. Statist.*, in the press.
Forbes, A. F. (1971). Non-parametric methods of estimating the survivor function. *The Statistician*, **20**, 27–52.
Fraser, D. A. S. (1968). *The Structure of Inference.* New York: Wiley.
Gart, J. J. (1970). Point and interval estimation of the common odds ratio in the combination of 2×2 tables with fixed marginals. *Biometrika*, **57**, 471–475.
Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skan. Aktuarietidskr.*, **39**, 125–153.
Harris, T. E., Meier, P. and Tukey, J. W. (1950). Timing of the distribution of events between observations. *Hum. Biol.*, **22**, 249–270.
Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters (with Discussion). *J. R. Statist. Soc.* B, **32**, 175–208.
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.
Nelson, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.*, **1**, 27–52.
Zelen, M. (1971). The analysis of several 2×2 contingency tables. *Biometrika*, **58**, 129–137.