Analysis of Survival Data under the Proportional Hazards Model
Author(s): N. E. Breslow
Source: *International Statistical Review / Revue Internationale de Statistique*, Vol. 43, No. 1 (Apr., 1975), pp. 45-57
Published by: International Statistical Institute (ISI)
Stable URL: https://www.jstor.org/stable/1402659
Accessed: 24-07-2018 13:42 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

International Statistical Institute (ISI) is collaborating with JSTOR to digitize, preserve and extend access to *International Statistical Review / Revue Internationale de Statistique*

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

Int. Stat. Rev., Vol. 43, No. 1, 1975, pp. 45–58/*Longman Group Ltd/Printed in Great Britain*

Analysis of Survival Data under the Proportional Hazards Model [1]

N. E. Breslow [2]
International Agency for Research on Cancer, Lyon, France

Summary

Methodology is reviewed for the statistical analysis of censored survival data which arise from a model in which the factors under investigation act multiplicatively on the hazard function of an underlying non-parametric survival distribution. This flexible approach provides computationally feasible solutions to the following problems: (i) one-sample problem (relative death rate); (ii) multi-sample problem; (iii) regression with continuous covariates; (iv) regression in matched-pair designs; (v) evaluation of changes in treatment or prognostic status (time dependent covariates). For the multi-sample problem with stratification, numerical results are presented contrasting maximum likelihood with simple chi-square analyses. While several of the methods described have been used on an *ad hoc* basis for many years, study of their common theoretical underpinnings has commenced only recently.

1. Introduction

This paper reviews a general methodology for the statistical analysis of survival or response time data such as accrue from the long-term follow-up of patients with chronic disease. The theoretical basis for this methodology is the assumption that prognostic or treatment factors under investigation have multiplicative effects on the hazard (instantaneous death rate) function of an underlying survival distribution. No particular parametric form is specified for this distribution.

The methods described are most appropriate when interest is focused on the actual duration of survival as a response variable. This will often be the case with chronic diseases of adulthood. If primary concern is instead in whether or not a patient survives a specified length of time, recourse is better made to related methods for dichotomous response variables [Bishop (1969), Cox (1970), Haberman (1973), Mantel (1966)].

An important feature of the methods considered is that they readily incorporate survival data which are "arbitrarily" right-censored. Censoring may be due to limitations on the observation period imposed by loss to follow-up, death from causes other than that under investigation, or survival to the study's end (withdrawal). It is assumed thoughout that these three causes of censorship operate independently of the cause(s) of death, and no distinction is made between the three. However, the reader is warned that, while commonly made, this assumption may not always be applicable [Chiang (1961), Mantel (1966)]. Care must be taken especially when analyzing mortality due solely to a specific cause.

2. Single Sample Comparison with a Life Table Population: the Relative Death Rate

A frequent goal of chronic disease studies is to compare the survival of a defined group of patients to that of a population of "normal" individuals having the same demographic (age, sex, race, etc.) characteristics. The "relative survival rate" has been used for this purpose

[1] Paper presented to the session on statistical aspects of chronic disease, 8th International Biometric Conference, Constanta, Romania, August 1974.

[2] Now at the University of Washington, Seattle.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

in cancer end results work by Ederer (1961), Cutler (1963) and Axtell (1969). However, the proportional hazards (PH) model suggests the use of the alternative "relative death rate", as described by Oleinick and Mantel (1970). Essentially the same statistic has long been used by epidemiologists [Doll, Morgan and Speizer (1970)] under the guise of indirect standardization; in fact Kirkpatrick (1962) has established optimal properties for the indirectly standardized mortality ratio (SMR) analogous to those noted below to hold for the relative death rate.

Suppose there are n patients of whom m die with (true) survival times t_i $(i = 1, \ldots, m)$ while $n-m$ are lost or withdrawn from observation with (censored) survival times

$$t_i \ (i = m+1, \ldots, n).$$

The survival distribution $F_i(t)$ for a person of the same demographic description as the ith patient may be obtained (see below) from published life tables. An appropriate formulation of the PH model for this situation is that the hazard function

$$\lambda_i(t) = F_i'(t)/(1-F_i(t)) \tag{1}$$

is multiplied by a constant θ, the relative death rate, which represents the excess mortality due to disease. This leads to the ln-likelihood function [1] for θ

$$L(\theta) = m \cdot \ln \theta + \theta \sum_{i=1}^{n} \ln (1-F_i(t_i)), \tag{2}$$

whence the maximum likelihood estimate (MLE) is

$$\hat{\theta} = m/-\sum_{i=1}^{n} \ln (1-F_i(t_i)) \tag{3}$$

while a chi-square statistic for testing $\theta = 1$ may be obtained as

$$[L'(1)]^2/-E(L''(1)) = \left[m + \sum_{i=1}^{n} \ln (1-F_i(t_i))\right]^2 \Big/ - \sum_{i=1}^{n} \ln (1-F_i(t_i)). \tag{4}$$

The denominator of (3) represents the number of deaths expected to occur in a demographically similar population, and the estimated relative death rate is simply the ratio of observed (O) to expected (E). The chi-square statistic (4) derived from the ln-likelihood is likewise simply $(O-E)^2/E$.

In practice the expected number of deaths E has often been calculated as follows. Each full year of observation for the ith patient contributes an amount equal to the conditional probability of death during that year, as obtained from the life table corresponding to his age at the beginning of the year, birth cohort, race, sex, etc. Partial years contribute only 1/2 (or the appropriate fractional part, if known) of the corresponding probability. From the viewpoint of the PH model it is clear that mid-year death rates [Chiang (1968)] should be used in preference to probabilities for this calculation, although these two quantities will be close when small.

Comparison of the relative death rates for two or more groups is best carried out on the ln scale since the large sample variance of ln $(\hat{\theta})$ is simply the reciprocal of the expected number of deaths. Thus the standardized difference of two such estimates, ln $(\hat{\theta}_1) =$ ln (O_1/E_1) and ln $(\hat{\theta}_2) =$ ln (O_2/E_2), is

$$\ln (O_1E_2/O_2E_1)/(E_1^{-1}+E_2^{-1})^{\frac{1}{2}}. \tag{5}$$

One would rarely expect the relative death rate to remain constant over many years of follow-up. Hence it is of interest to estimate θ for groups of years, or even single years, of follow-up and plot the resulting estimates against time. A curve which started high but

[1] Through this paper additive constants which depend on the observations but not the parameters are suppressed from formulas for the ln-likelihood.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

decreased towards unity could indicate that the long term survivors had been "cured". Appropriate smoothing of this curve could be accomplished by calculating θ for overlapping groups of follow-up years, e.g. 1–3, 2–4, 3–5, etc.

3. Comparison of Several Homogeneous Samples: Application of Standard Chi-square Techniques

In other follow-up studies of chronic disease patients, particularly in randomized clinical trials, one will be interested in a direct comparison of survival for different treatment or prognostic groups without reference to a life table population. The survival time for an individual in the ith of r different groups is considered as a random observation from a distribution $F_i(t)$. Under the PH model for this situation the hazard functions for each of these distributions are assumed proportional: $\lambda_i(t)/\lambda_j(t) \equiv C$ (constant for all t). Equivalently, the survival functions are assumed related by a power transform: $(1-F_i(t)) \equiv (1-F_j(t))^C$.

Mantel (1963, 1966) pointed out that standard methods for combining chi-square statistics from a series of independent contingency tables [Armitage (1966), Cochran (1954), Mantel and Haenszel (1959)] could be adapted to this situation and that the power function of the resulting test was constant for alternatives specified by the PH model. Peto and Peto (1972) showed that it was an "asymptotically efficient, rank invariant" test for this model, while Peto (1972*b*) proved that it also had optimal properties locally (in the region of the null hypothesis), provided that censorship patterns were equal in all groups. Another derivation was given by Cox (1972) in the context of his general regression model. Some authors [Crowley (1973), Thomas (1971)] have referred to the test as the "generalized Savage" test since, if there is no censorship, it reduces to Savage's test based on exponential ordered scores [Cox (1964), Savage (1956)]. However it is presented here from Mantel's original viewpoint since this makes most transparent its relationship to modern methods of contingency table analysis.

Suppose now that for n individuals in r treatment or prognostic groups there are exactly K distinct, ordered, true (uncensored) survival times $t_{(1)}<\ldots<t_{(K)}$. Consider the K corresponding $2\times r$ contingency tables ($k = 1, \ldots, K$).

A_{1k}	A_{2k}		A_{rk}	N_{1k}
B_{1k}	B_{2k}		B_{rk}	N_{2k},
M_{1k}	M_{2k}		M_{rk}	T_k

(6)

where A_{ik} is the number of deaths at $t_{(k)}$ among patients in the ith group ($i = 1, \ldots, r$) while $M_{ik} = A_{ik}+B_{ik}$ is the number in that group alive and under observation (not lost or withdrawn) just prior to $t_{(k)}$. If there are no ties among the true survival times all the N_{1k} will equal 1 and K will equal the number of deaths. A test for the homogeneity of the r survival curves is obtained by calculating the expectation $\boldsymbol{E}_k$ and covariance matrix $\boldsymbol{V}_k$ of the observation vector $\boldsymbol{O}_k = (A_{1k}, \ldots, A_{rk})^T$ in each table under the hypothesis of fixed marginals [Mantel and Haenszel (1959)] and then summing these quantities over the K tables. The summary test statistic, which may be referred to tables of chi-square with $r-1$ degrees of freedom, is

$$\chi^2_{r-1} = (\boldsymbol{O}-\boldsymbol{E})^T \boldsymbol{V}^- (\boldsymbol{O}-\boldsymbol{E}) \tag{7}$$

where $\boldsymbol{O} = \Sigma_k \boldsymbol{O}_k$, $\boldsymbol{E} = \Sigma_k \boldsymbol{E}_k$, $\boldsymbol{V} = \Sigma_k \boldsymbol{V}_k$, and $\boldsymbol{V}^-$ is a generalized inverse. In practice one drops all terms from $\boldsymbol{O}$, $\boldsymbol{E}$, and $\boldsymbol{V}$ corresponding to the last (rth) group, and performs the same calculations (normal matrix inversion) for the resulting $r-1$ dimensional problem. When $r = 2$ a continuity correction of $\frac{1}{2}$ should be used, as is done routinely in the numerical examples below.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

A conservative approximation to this test statistic is obtained from the usual formula $\Sigma_i (O_i - E_i)^2/E_i$, where the sum is over the r components of the summary observation and expectation vectors $\boldsymbol{O}$ and $\boldsymbol{E}$ [Peto and Pike (1973)]. Use of the approximation in large samples has been shown to result in little loss of power unless there are marked differences in the withdrawal patterns for the r groups or many ties [Crowley and Breslow (1974)].

This procedure is easily modified to adjust for heterogeneity in the comparison groups by means of stratification. The study population is divided into several relatively homogeneous strata on the basis, for example, of prognostic variables and a separate comparison of the r groups is made within each one. For an overall test, the summary statistics $\boldsymbol{O}$, $\boldsymbol{E}$ and $\boldsymbol{V}$ calculated for each stratum are pooled by simple addition and then treated as before.

The ratio O/E, which was used to estimate the relative death rate in the one sample problem, was proposed by Pike (1972) for that same purpose with multiple samples. While very useful as a descriptive statistic, O/E unfortunately does not estimate any well-defined parameter in this case. With two samples, for example, the fact that E_1 is calculated using data from both samples means that O_1/E_1 is biased towards unity as an estimate of the death rate ratio $\theta = \lambda_1(t)/\lambda_2(t)$. A more appropriate estimate of θ in this case is the ratio of O/E for the two samples, namely $O_1/E_1 \div O_2/E_2$.

The results of Crowley (1973) may be used to assess the adequacy of $O_1/E_1 \div O_2/E_2$ as an estimate of θ in large samples. If there is no censorship, then O_1/E_1 converges to the quantity

$$\alpha(\theta, R) = \left[\int_0^1 \frac{dx}{1+Rx^{(\frac{1}{\theta}-1)}} + \frac{1}{\theta}\int_0^1 \frac{dx}{1+R^{-1}x^{(1-\frac{1}{\theta})}}\right]^{-1}, \tag{8}$$

where R is the ratio of the second to the first sample size. Consequently $O_1/E_1 \div O_2/E_2$ will converge to $\alpha(\theta, R) \div \alpha(\theta^{-1}, R^{-1})$. A brief tabulation of this quantity (Table I) indicates that the bias is minimal for $\frac{1}{2} \leqq \theta \leqq 2$ but that extreme values of θ may be underestimated. Similar conclusions are drawn from numerical comparisons with the MLE, made in cases involving censorship, stratification and more than two comparison groups (Table III).

Table I. *Expected values (asymptotic) of $O_1/E_1 \div O_2/E_2$ as a measure of the relative death rate for the two sample problem without censorship*

θ = Ratio of hazards for sample 1 to sample 2	R = Ratio of size of sample 2 to sample 1										
	0·00	0·05	0.10	0·20	0·50	1·00	2·00	5·00	10·00	20·00	∞
1·0	1·0	1·0	1·0	1·0	1·0	1.0	1·0	1·0	1·0	1·0	1·0
1·2	1·2	1·200	1·199	1·199	1·198	1·198	1·199	1·199	1·200	1·200	1·2
1·5	1·5	1·492	1·487	1·482	1·478	1·479	1·483	1·491	1·495	1·497	1·5
2·0	2·0	1·930	1·905	1·883	1·873	1·886	1·912	1·950	1·972	1·985	2·0
5·0	5·0	3·288	3·143	3·056	3·092	3·277	3·597	4·115	4·455	4·692	5·0

An alternative estimate of θ for two groups is provided by the Mantel-Haenszel combined relative risk formula (1959) applied to the series of tables (6). This is a close approximation to the MLE when θ is near unity [Clayton (1974)]. The iterative MLE itself will be discussed below in terms of the general regression model. Other non-iterative estimates have been studied in the context of contingency table analysis [Gart (1970)].

These methods are illustrated with preliminary data from a trial of maintenance chemotherapy for children with acute lymphocytic leukemia [Miller *et al.* (1974), Breslow (1974)]. The response variable for this trial was the duration of remission after successful completion of induction chemotherapy. Treatment group (regimen) 5 received the standard maintenance

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

drugs; the others received in addition actinomycin-D (regimens 1, 2, 4) and/or nitrogen mustard (regimens 1, 3, 4). Stratification was according to the diagnostic white blood count (WBC), the major prognostic indicator.

Table II*a* shows the numbers of patients entered and the number of relapses by treatment and WBC. Table II*b* examines the joint effects of these two variables, the data being split into $r = 15$ groups for calculation of E and V (not shown). The marginal totals of expected values are precisely those obtained when an analysis is performed ignoring the alternate variable; thus, comparing the three WBC strata while ignoring treatment, the expected numbers of relapses are 72·63, 59·92 and 48·45, etc. Likewise variances and covariances for the summed (marginal) observations may be obtained from appropriate terms in the 15-dimensional covariance matrix. This leads to significance tests for the effects of treatment or WBC, unadjusted for the other factor. For example, comparison of observed (100) and expected (113·04) numbers of relapses in the three regimens employing actinomycin-D yields a $\chi^2_1 = 3{\cdot}68$ using the $\Sigma\,(O-E)^2/E$ approximation or $\chi^2_1 = 3{\cdot}76$ without the approximation. The multiplicative effect of this drug on the hazard function estimated by $O_1/E_1 \div O_2/E_2$ is $100/113{\cdot}0 \div 81/68{\cdot}0 = 0{\cdot}74$, precisely the value obtained by maximum likelihood [Breslow (1974), Table 2, line 4].

From Table II*b* it is clear that WBC plays a major prognostic role, with remission durations shortening as WBC increases in each of the four treatment regimens. Table II*c* shows treatment effects adjusted for WBC, obtained by carrying out a separate analysis in each of the three strata and then pooling summary statistics. The expected number of relapses among regimens employing ACT-D is 110·96 after adjustment, leading to an adjusted chi-square of $\chi^2_1 = 2{\cdot}55$ using the $\Sigma\,(O-E)^2/E$ approximation or $\chi^2_1 = 2{\cdot}64$ without. Adjusting for WBC lowers the significance of the drug effect since (Table II*a*) a lesser proportion (28 per cent) of patients in regimens 1, 2, and 4 have a WBC over 20,000 than for regimens 3 and 5 (37 per cent). Table II*d* presents a similar analysis for the effect of WBC, which is strong and little affected by adjustment for treatment.

Close examination of Table II*c* reveals a small interaction effect between treatment and WBC: the control regimen is inferior to all four treatment regimens for WBC < 20,000; it is apparently superior for patients with the least favourable prognosis. While statistically insignificant with these data (Table V), this interaction was confirmed by a more elaborate analysis of the final data set [Miller *et al.* (1974)] and interpreted to mean that conventional maintenance therapy may be of marginal benefit to high risk patients.

4. Accounting for Continuous Covariates: Cox's Regression Model

The demonstration by Cox (1972) that a regression analysis could be carried out with the PH model has stimulated much further work with this model. His approach to regression is embodied in the equation

$$\lambda\,(t;\ z) = e^{\beta' z}\lambda_0\,(t) \tag{9}$$

where $\lambda\,(t;\ z)$ is the hazard function for an individual with a p-vector z of covariates (which may include both treatment and prognostic factors), β is the p-vector of regression coefficients, and $\lambda_0\,(t)$ is the hazard function of the underlying survival distribution F. An attractive by-product of this approach is a non-parametric estimate of F which generalizes the familiar product-limit estimate [Kaplan and Meier (1958), Breslow and Crowley (1974)] to heterogeneous samples.

In comparison with procedures for grouped data, the regression approach does not lend itself so easily to tied data since F is implicitly assumed by (9) to be continuous. For this reason and also to estimate F, Cox (1972) proposed and used in calculations a discrete analog to (9) in which conditional probability masses are placed on each of the distinct, ordered,

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

Table II. *Analysis of effects of treatment group (RX) and diagnostic white blood count (WBC) on the remission duration of 268 children with acute lymphocytic leukaemia*

(a) *Number of patients (N) and observed number of relapses (O), by RX and WBC*

	RX											
	1		2		3		4		5		Totals	
WBC	*N*	*O*	*N*	*O*	*N*	*O*	*N*	*O*	*N*	*O*	*N*	*O*
0–	19	8	13	7	20	10	19	10	16	8	87	43
5,000–	14	7	22	13	21	15	22	18	16	3	95	66
20,000–	13	11	17	14	23	21	13	12	20	14	86	72
Totals	46	26	52	34	64	46	54	40	52	35	268	181

(b) *Joint effects of RX and WBC: expected numbers of relapses (E) and ratio of observed to expected (O/E)*

	RX											
	1		2		3		4		5		Totals	
WBC	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*
0–	16·69	0·48	12·41	0·56	15·66	0·64	15·94	0·63	11·93	0·67	72·63	0·59
5,000–	10·70	0·65	17·01	0·76	11·54	1·30	13·91	1·29	6·78	1·92	59·92	1·10
20,000–	8·94	1·23	10·44	1·34	9·04	2·32	6·97	1·72	13·05	1·07	48·45	1·49
Totals	36·33	0·72	39·80	0·85	36·24	1·27	36·81	1·09	31·76	1·10	181·00	1·00

(c) *Effects of RX adjusted for WBC: expected numbers of relapses (E) and ratio of observed to Expected (O/E)*

	RX											
	1		2		3		4		5		Totals	
WBC	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*
0–	9·67	0·83	8·42	0.83	9·08	1·10	8·99	1·11	6·84	1·17	43·00	1·00
5,000–	11·75	0·60	18·16	0·72	13·44	1·12	14·33	1·20	7·72	1·68	66·00	1·00
20,000–	13·14	0·84	15·68	0·89	13·88	1·51	10·22	1·17	19·08	0·73	72·00	1·00
Totals	34·56	0·75	42·26	0·80	36·39	1·26	34·14	1·17	33·65	1·04	181·00	1·00

(d) *Effects of WBC adjusted for RX: expected number of relapses (E) and ratio of observed to expected (O/E)*

	RX											
	1		2		3		4		5		Totals	
WBC	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*	*E*	*O/E*
0–	12·65	0·65	9·95	0·70	19·93	0·50	18·10	0·55	12·82	0·62	73·46	0·59
5,000–	7·07	0·99	14·53	0·89	14·72	1·02	14·61	1·23	7·73	1·68	58·67	1·12
20,000–	6·27	1·75	9·52	1·47	11·35	1·85	7·29	1·65	14·45	0·97	48·87	1·47
Totals	26·00	1·00	34·00	1·00	46·00	1·00	40·00	1·00	35·00	1·00	181·00	1·00

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

uncensored survival times $t_{(k)}$ ($k = 1, \ldots, K$). These are related through a linear logistic equation [Cox (1970), p. 18].

$$\text{logit } p_k(z) = \boldsymbol{\beta}'z + \text{logit } p_k, \tag{10}$$

where $p_k(z)$ is the conditional probability that a patient with covariables z dies at $t_{(k)}$, given that he was alive and under observation at $t_{(k)}-0$. The underlying survival distribution F has corresponding quantities p_k. Unfortunately, maximum likelihood calculations based on the model (10) are cumbersome when ties are heavy. Iteration is required at each $t_{(k)}$ in order to estimate the p_k.

Another approach to inferences about $\boldsymbol{\beta}$, which is more in the spirit of the continuous model (9), has been proposed by Peto (1972*a*) (his "real probability") and Kalbfleisch and Prentice (1973). Here a marginal likelihood function for $\boldsymbol{\beta}$ is obtained by calculating the probability under (9) of all possible rank orderings of the n times of death consistent with the observed pattern of ties and censorship. In order to estimate F, Kalbfleisch and Prentice suggest an alternative discrete model which is specified by the equation

$$\ln(1-p_k(z)) = e^{\boldsymbol{\beta}'z} \ln(1-p_k) \tag{11}$$

for the conditional probability masses defined above. This arises from (9) upon partitioning of the continuous sample space. With (11) iterative calculations for estimation of the p_k are required only at those $t_{(k)}$ which represent multiple deaths.

While one of these two approaches should undoubtedly be used with heavily tied data, I have found [Breslow (1972), (1974)] a simpler set of equations adequate for dealing with large sets of essentially continuous data. These are obtained by assuming the survival distribution F to have a hazard function which is constant $\lambda_0(t) = \lambda_k = \exp(\alpha_k)$ for t in the interval $(t_{(k-1)}, t_{(k)}]$ between each pair of uncensored survival times ($k = 1, \ldots, K$). If each censored survival time is adjusted to have occurred at the preceding uncensored time, one arrives at a joint ln-likelihood function for λ_0 (in terms of the α_k) and $\boldsymbol{\beta}$:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{k=1}^{K} (\alpha_k m_k + \boldsymbol{\beta}'s_k - e^{\alpha_k}(t_{(k)}-t_{(k-1)}) \sum_{j \in R_k} e^{\boldsymbol{\beta}'z_j}). \tag{12}$$

Here m_k is the number of true deaths occurring at $t_{(k)}$, s_k is the sum of the covariates for these m_k individuals, and R_k is the "risk set" of all individuals alive and under observation at $t_{(k)}-0$. Maximizing (12) with respect to $\boldsymbol{\alpha}$ leads to the simple formula

$$\ln(1-\hat{F}(t_{(k)})) = -\sum_{i=1}^{k} m_i / \sum_{j \in R_i} e^{\boldsymbol{\beta}'z_j} \tag{13}$$

as the MLE (in terms of $\boldsymbol{\beta}$) of F at $t_{(k)}$ [Oakes (1972)]. Inferences regarding $\boldsymbol{\beta}$ are made from the "maximum" ln-likelihood function

$$L(\boldsymbol{\beta}) = \sup_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{k=1}^{K} (\boldsymbol{\beta}'s_k - m_k \ln \sum_{j \in R_k} \exp(\boldsymbol{\beta}'z_j)). \tag{14}$$

Equation (14), which corresponds to Peto's (1972*a*) "rough probability", requires but a single summation of the terms $\exp(\boldsymbol{\beta}'z_j)$ for individuals in the risk set at $t_{(k)}-0$. The more exact approaches discussed above required multiple summations unless $m_k = 1$, in which case all three give the same contribution to the ln-likelihood at $t_{(k)}$.

A detailed analysis based on (13) and (14) of the data on 268 leukaemic children, treating WBC and age at diagnosis as continuous covariates, has been published [Breslow (1974)] and will not be repeated here. Instead the grouped data will be re-analyzed in the maximum likelihood framework for comparison with the simpler chi-square procedures. This means that the covariate vector z is composed of binary variables which identify the treatment and/or WBC stratum for each patient. Care must be taken to eliminate functionally dependent components, so as to avoid singular information matrices during iterative solution of the likelihood equations.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

The joint effects of treatment and WBC were estimated by taking for z a 14-dimensional vector, corresponding to the 15 treatment × WBC categories minus one. The category eliminated ($z = \mathbf{0}$ for individuals in this category) was regimen 5, WBC < 5,000, which thus serves as a standard. Estimated regression coefficients β_i for the 14 other categories were exponentiated to yield the multiplicative effects shown in Table III.

Main effects for treatment and WBC, adjusted for each other, were determined as regression coefficients corresponding to a 6-dimensional binary covariate vector in which the first 4 components identified the treatment (regimen 5 as standard) and the remaining 2 the WBC stratum (<5,000 as standard). After conversion to multiplicative effects by exponentiation, these are presented in the margins of Table III. Multiplication of the two corresponding marginal effects yields an estimated effect for each cell, which may be divided into that cell's

Table III. *Joint and main effects of treatment (RX) and white blood count (WBC), estimated by maximum likelihood and O/E (in parenthesis)*

	RX					
WBC	1	2	3	4	5	Main effect (WBC)
0–	0·71	0·82	0·94	0·93	1·0	1·0
	(0·71)	(0·84)	(0·95)	(0·94)	(1·0)	(1·0)
5,000–	0·98	1·13	1·99	1·95	2·94	1·91
	(0·98)	(1·14)	(1·94)	(1·93)	(2·86)	(1·92)
20,000–	1·85	2·00	3·55	2·61	1·60	2·58
	(1·83)	(2·00)	(3·46)	(2·57)	(1·60)	(2·52)
Main effects	0·72	0·78	1·27	1·11	1·0	—
(RX)	(0·72)	(0·77)	(1·22)	(1·13)	(1·0)	—

Table IV. "*Interaction*" *effects between RX and WBC*

	RX				
WBC	1	2	3	4	5
0–	0·99	1·05	0·74	0·84	1·0
5,000–	0·71	0·76	0·82	0·92	1·54
20,000–	1·00	0·99	1·08	0·91	0·62

joint effect to measure "interactions" (Table IV). For example the estimated effect for regimen 2, WBC over 20,000 is $0{\cdot}78 \times 2{\cdot}58 = 2{\cdot}01$ compared with 2·00 observed. Not surprisingly, the largest interactions are found with the two cells for regimen 5 and WBC over 5,000.

Table III also compares these MLE's to the simpler estimates in Table II, after appropriate standardization. Thus the O/E ratios in Table II*b* were all divided by the ratio 0·67 for regimen 5, WBC < 5,000. Main effects for treatment and WBC were obtained from Tables II*c* and II*d*, respectively. These show remarkably good agreement with the MLE's although extreme effects tend to be underestimated slightly.

A comparison of likelihood ratio statistics with those obtained in the previous chi-square analysis is made in Table V. The "combination" chi-square is that calculated using the summary covariance matrix V as in (7). The statistics for testing the adjusted effects of treatment and WBC using the chi-square analysis are obtained from Table II*c* and II*d*; the others are from Table II*b*. Approximate chi-square tests for combined main effects and interactions could be obtained by subtraction.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

Table V. "*Anova*" *of treatment* (*RX*) *and white count* (*WBC*) *effects*

Source	d.f.	Ln-likelihood ratio	Chi-square	
			Combination	$\Sigma (O-E)^2/E$
WBC and RX (main effects)	6	32·92	—	—
WBC alone	2	24·95	24·55	24·16
RX adjusted for WBC	4	7·97	7·58	7·33
RX alone	4	7·24	7·15	7·03
WBC adjusted for RX	2	25·68	26·13	24·49
WBC × RX (interaction)	8	7·94	—	—
Total	14	40·86	45·09	43·81

5. The Analysis of Survival Data in Matched Pairs

An interesting application of the PH model to the regression analysis of survival times occurring in matched pairs has recently been made by Holt and Prentice (1974). In their approach, which is particularly relevant to twin studies, it is assumed that the underlying survival distribution may vary from pair to pair, but that the covariates continue to act multiplicatively on the hazard functions. Thus the model equation becomes

$$\lambda_i(t, z) = e^{\beta' z}\lambda_{0i}(t) \tag{15}$$

for the hazard function of a member of the ith pair having covariates z. A marginal likelihood for β is obtained by considering the order of the survival times in each pair. In order to incorporate censoring, it is necessary to assume that the withdrawal times are identical for pair mates. While this will be true for twin studies and for matched pair mates which come under observation at the same time, it will not hold if pairs are formed retrospectively from a pool of cases having different observation periods. Nor can censorship due to losses or competing causes of death be incorporated without additional assumptions.

Suppose then that t_{i1} and t_{i2} denote the true survival times of the ith set of pair mates having covariates z_{i1} and z_{i2}, with difference $d_i = z_{i1} - z_{i2}$, while t_i^0 is the common limit on observation. In order to ascertain the rank order at least one observation on survival must be uncensored. In this case there is a (conditional) contribution to the likelihood of the form

$$P(t_{i1} < t_{i2} \mid t_{i1} < t_i^0 \text{ or } t_{i2} < t_i^0) = \{1 + \exp(-\beta' d_i)\}^{-1}, \tag{16}$$

so that the ln-likelihood function itself is given by

$$L(\beta) = -\sum_{i=1}^{K} \ln\{1 + \exp(\varepsilon_i \beta' d_i)\}, \tag{17}$$

where the sum is over the K uncensored or singly censored pairs and $\varepsilon_i = +1$ or -1 according as $t_{i2} < t_{i1}$ or $t_{i1} < t_{i2}$. Since only ranks within each pair are considered, ties should be infrequent, in which case they can be ignored.

If the covariate vector z simply indexes which of two treatment groups each twin is in, then the model (15) specifies that the survival curve for the treated twin is that for the untreated twin raised to a constant power $\theta = \exp(\beta)$. In this case inferences based on (17) reduce to the familiar sign method [Armitage (1959)]: the MLE of $\hat{\theta}$ is the ratio of the number of pairs in which the treated twin dies first to the number where the untreated twin dies first while the hypothesis $\theta = 1$ may be tested by the usual (McNamara) test for equality of proportions in matched samples. It has long been recognized that, by considering each pair as a separate stratum, the Mantel-Haenszel (1959) chi-square procedures yield precisely the same estimate and test for this problem.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

While extensions of this method for the analysis of data collected in a randomized blocks design could be made, these would be even more limited in applicability by the requirement that all patients in a block have the same censoring time. Further work is needed to determine how critical this assumption really is.

6. Accounting for Changes in Treatment or Prognostic Status: the use of Time Dependent Covariates

So far in this paper it has been assumed, albeit implicitly, that a patient's treatment group and his values on prognostic variables are recorded once, at that point in time (first symptoms, diagnosis, hospital admission, treatment, etc.) from which survival is measured for all patients. The effects of these fixed values on subsequent survival are examined, but no account is taken of changes in a patient's status which may occur during the follow-up period. With many chronic disease studies such changes are bound to occur and are in fact of principal interest. As shown now for a simple and particularly interesting example, such changes may prove to be easily incorporated into a statistical analysis based on the PH model.

Suppose that a single change in a patient's treatment status, say from untreated to treated, may occur in the course of follow-up. This is the case with heart transplant patients who are kept under observation until a suitable donor can be found, at which time they receive their transplant. A simple comparison of the survival times of treated with untreated patients fails to account for the fact that the former must live at least long enough to be treated [Gail (1972)]. Several authors [Crowley (1973), Mantel and Byar (1974), Turnbull, Brown and Hu (1973)] have more or less explicitly pointed out that a valid approach to this problem may be based on a variation of the PH model whereby the hazard function for a patient treated x time units after entry into observation is assumed to satisfy

$$\lambda(t;\, x) = \begin{cases} \lambda_0(t), & t < x \\ e^{\beta}\lambda_0(t), & t \geqq x. \end{cases} \tag{18}$$

Thus e^{β} represents the multiplicative treatment effect while $\lambda_0(t)$ is the hazard function for an untreated patient.

In order to explore the methodological consequences of (18), let $t_{(1)} < \ldots < t_{(K)}$ denote once again the ordered, distinct, uncensored times of death. Suppose n_k patients are at risk at $t_{(k)} - 0$, of whom n_{k1} have received their treatment and $n_{k2} = n_k - n_{k1}$ have not; while of m_k deaths occurring at $t_{(k)}$, d_{k1} are among untreated and $d_{k2} = m_k - d_{k1}$ among treated patients ($k = 1, \ldots, K$). An extension of the arguments of [Breslow (1972)] shows that an approximate "maximum" ln-likelihood function for β in this circumstance is

$$L(\beta) = \sum_{k=1}^{K} (\beta d_{k_2} - m_k \ln(n_{k_1} + e^{\beta} n_{k_2})), \tag{19}$$

which is precisely the same as (14) for the case of two independent samples of treated and untreated patients followed from diagnosis. Thus, although in the present case a patient may move from one treatment group to another as time progresses, nevertheless the statistical procedures remain the same. In particular likelihood inferences about β may be made directly from (19), while the chi-square procedures may be based on the series of 2×2 tables

$$(i = 1, \ldots, K)$$

d_{i1}	d_{i2}
$n_{i1} - d_{i1}$	$n_{i2} - d_{i2}$

. (20)

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

These latter allow adjustment for (fixed) prognostic factors by means of stratification. However, the $\Sigma (O-E)^2/E$ approximations should not be used with time dependent treatment status since this is a case in which the proportions of patients at risk in each group will change markedly with time [Crowley and Breslow (1974)].

Extensions of this approach can clearly be made to examine the effect of membership in any one of a finite number of transient states provided that risk of death is determined solely by the current state of each patient and not by his past history. A limited amount of dependence on previous transitions could be obtained by considering each possible path through the transient states as a separate "historical" state [Mantel and Byar (1974)].

From the point of view of the general regression model (9), (18) is obtained by the use of a single time-dependent covariate z defined by

$$z(t) = \varepsilon(t-x), \tag{21}$$

where $\varepsilon(u) = 1$ or 0 as $u \geqq 0$ or $u<0$. The use of such covariates was suggested in the original paper by Cox (1972) but has been questioned by later authors [Kalbfleisch and Prentice (1973)]. The fact that the likelihood function (19) leads to valid inferences, as shown by Crowley (1973), would seem to justify its use in this particular case. It would be of interest to know whether inferences based (9) were generally valid in the case of time-dependent covariates, since this would permit one to estimate simultaneously their effects and the effects of fixed covariates. With the chi-square approach, this would mean stratifying on time-dependent as well as fixed covariates for the estimation of "adjusted" effects.

References

Armitage, P. (1959). The comparison of survival curves. *Journal of the Royal Statistical Society* (A), **122**, 279–300.

Armitage, P. (1966). The χ^2 test for heterogeneity of proportions after adjustment for stratification. *Journal of the Royal Statistical Society* (B), **28**, 150–163.

Axtell, L. (1969). Life table program: observed and relative rates. Methodological note no. 24, End Results Section, National Cancer Institute, Bethesda, Maryland.

Bishop, Y. M. M. (1969). Full contingency tables, logits, and split contingency tables. *Biometrics*, **25**, 119–128.

Breslow, N. (1972). Contribution to the discussion on the paper by D. R. Cox, Regression Models and Life Tables. *Journal of the Royal Statistical Society* (B), **34**, 216–217.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89–100.

Breslow, N., Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Annals of Statistics*, **2**, 437–453.

Chiang, C. L. (1961). On the probability of death from specific causes in the presence of competing risks. *Proceedings of the 4th Berkeley Symposium*, **4**, 169–180. University of California Press, Berkeley-Los Angeles.

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

Clayton, D. G. (1971). Multiplicative models in 2-dimensional tables of rates. (Submitted for publication.)

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

Cox, D. R. (1964). Some applications of exponential ordered scores. *Journal of the Royal Statistical Society* (B), **26**, 103–110.

Cox, D. R. (1970). *The Analysis of Binary Data*. Methuen, London.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* (B), **34**, 187–220.

Crowley, J. (1973). Non-parametric analysis of censored survival data with distribution theory for the k-sample generalized Savage statistic. Ph.D. Thesis. University of Washington.

Crowley, J., Breslow, N. (1974). Remarks on the conservatism of $\Sigma (O-E)^2/E$ in survival data. (Submitted for publication.)

Cutler, S. J., Axtell, L. (1963). Partitioning of a patient population with respect to different mortality risks. *Journal of the American Statistical Association*, **58**, 701–712.

Doll, R., Morgan, L. G., Speizer, F. E. (1970). Cancers of the lung and nasal sinuses in nickel workers. *British Journal of Cancer*, **25**, 623–632.

Ederer, F., Axtell, L., Cutler, S. J. (1961). The relative survival rate: a statistical methodology. *National Cancer Institute Monograph Number*, **6**, 101–121. United States Government Printing Office, Washington.

Gail, M. H. (1972). Does cardiac transplantation prolong life? A reassessment. *Annals of Internal Medicine*, **76**, 815–817.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

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.
Haberman, S. J. (1973). Log-linear models for frequency data: sufficient statistics and likelihood equations. *Annals of Statistics*, **1**, 617–632.
Holt, J. D., Prentice, R. L. (1974). Survival analyses in twin studies and matched pair experiments. *Biometrika*, **61**, 17–30.
Kalbfleisch, J. D., Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, **60**, 267–278.
Kaplan, E. L., Meier, P. (1958). Non-parametric estimation from incomplete observation. *Journal of the American Statistical Association*, **53**, 457–481.
Kirkpatrick, S. J. (1962). Occupational mortality indices. *Population Studies*, **16**, 175–182.
Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, **58**, 690–700.
Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**, 163–170.
Mantel, N. (1966). Models for complex contingency tables and polychotomous dosage response curves. *Biometrics*, **22**, 83–95.
Mantel, N., Byar, D. P. (1974). Evaluation of response-time data involving transient states: an illustration using heart transplant data. *Journal of the American Statistical Association*, **69**, 81–86.
Mantel, N., Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**, 719–748.
Miller, D. R., Sonley, M., Karon, M., Breslow, N., Hammond, D. (1974). Additive therapy in the maintenance of remission in acute lymphoblastic leukemia of childhood: the effect of prognostic factors. *Cancer*, **34**, 508–517.
Oakes, D. (1972). Contribution to the discussion on the paper by D. R. Cox, Regression Models and Life Tables. *Journal of the Royal Statistical Society* (B), **34**, 208.
Oleinick, A., Mantel, N. (1970). Family studies in systemic lupus erythematosus – II. *Journal of Chronic Diseases*, **22**, 617–625.
Peto, R. (1972*a*). Contribution to the discussion on the paper by D. R. Cox, Regression Models and Life Tables. *Journal of the Royal Statistical Society* (B), **34**, 205–207.
Peto, R. (1972*b*). Rank tests of maximum power against Lehmann type alternatives. *Biometrika*, **59**, 472–474.
Peto, R., Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society* (A), **135**, 185–206.
Peto, R., Pike, M. C. (1973). Conservatism of the approximation $\Sigma (O-E)^2/E$ in the logrank test for survival data or tumor incidence data. *Biometrics*, **29**, 579–583.
Pike, M. C. (1972). Contribution to the discussion on the paper by Peto, R. and Peto, J., Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society* (A), **135**, 201–203.
Savage, I. R. (1956). Contributions to the theory of rank order statistics: the two sample case. *Annals of Mathematical Statistics*, **27**, 590–615.
Thomas, D. R. (1971). On the asymptotic normality of a generalized Savage statistic for comparing two arbitrarily censored samples. Technical Report. Department of Statistics, Oregon State University.
Turnbull, B. W., Brown, B. W., Jr., Hu, M. (1973). Survivorship analysis of heart transplant data. *Journal of the American Statistical Association*, **69**, 74–80.

Acknowledgement

Several helpful comments on earlier drafts of this paper were made by N. Mantel and R. Peto.

Résumé

Cet article présente une revue des méthodes utilisées pour l'analyse statistique d'observations censurées de durée de vie, qui résultent d'un modèle où les facteurs étudiés agissent multiplicativement sur la fonction de hasard d'une distribution non paramétrique sous-jacente du nombre de survivants. Cette approche flexible fournit des solutions aisément calculables aux problèmes suivants: (1) problème portant sur un seul échantillon (relatif au taux de décès); (2) problème à plusieurs échantillons; (3) régression à variables continues; (4) régression dans les plans avec observation par paire; (5) évaluation de la modification du traitement ou du pronostic (covariables dépendant du temps). Dans le problème à plusieurs échantillons avec stratification, des exemples numériques sont présentés, qui confrontent le maximum de vraisemblance aux méthodes plus simples d'analyse de khi-carré. Bien que plusieurs des méthodes décrites aient été utilisées sur une base convenable pendant de nombreuses années, une étude de leur rattachement théorique commun n'a commencé que récemment.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms

Note added in Proof

A revised version of this paper, which arrived after submission to the printers, made the following additional points:

(i) The arguments leading to equations (16) and (17) are valid for t_i^0 equal to the minimum of two possibly distinct observation limits on the survival times of the two pair members, so that the hypothesis of equal censorship is unnecessary. The ln-likelihood (17) can also be obtained by summing individual contributions of the form (14) from each pair; such contributions are non-zero only in the case both pair members are "at risk" just prior to the time of the first true death in each pair. Following up this approach, extensions of the method for the analysis of data collected in randomized blocks or more general stratified designs can be made simply by adding together ln-likelihood (14) contributions calculated separately within each stratum.

(ii) In recent work, Crowley has shown that the arguments leading to (14) as the "maximum" ln-likelihood function apply as well in the case of time-dependent covariates as they do for fixed covariates. One simply evaluates the terms $s_k(t)$ and $z_j(t)$, occurring in the kth summand of (14), at the time of the kth death $t = t_{(k)}$. However, it remains to be seen if this approach can be justified also using other more rigorous arguments.

This content downloaded from 128.164.162.77 on Tue, 24 Jul 2018 13:42:56 UTC
All use subject to https://about.jstor.org/terms