ESTIMATION IN THE GENERAL MULTIPLICATIVE MODEL FOR SURVIVAL

TECHNICAL

NORMAN L. JOHNSON

AND

REGINA C. ELANDT-JOHNSON

JULY 1978

U. S. ARMY RESEARCH OFFICE
RESEARCH TRIANGLE PARK, NORTH CAROLINA

DAAG29-77-C-0035

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27514

ESTIMATION IN THE GENERAL MULTIPLICATIVE MODEL FOR SURVIVAL

Norman L. Johnson*, Department of Statistics
and
Regina C. Elandt-Johnson**, Department of Biostatistics
University of North Carolina at Chapel Hill

SUMMARY

The general multiplicative model (formula (2)) represents the hazard function as the product of an "underlying" hazard rate, $\lambda(t)$, of unspecified form and a certain function of known form, $g(\underset{\sim}{z}; \underset{\sim}{\beta})$, where $\underset{\sim}{z}$ is a vector of concomitant variables, and $\underset{\sim}{\beta}$ is a vector of unknown parameters. Assuming that $\lambda(t)$ can be approximated by a constant between any two consecutive failures, the general forms of likelihood function are derived (formulae (6) and (7), or (9) and (11)). The likelihood utilizes the available information on the time of exposure to risk of each individual (until failure or withdrawal). Special cases, when the z's do not depend on t are discussed in some detail (Section 7). Multiple failures are handled in a simple manner — no ordering of failures is required (Section 8). Estimation of empirical survival function when there are no covariates is discussed in Section 9. An example using heart transplant data, is given (for illustrative purpose only).

---

*Research supported by the Army Office of Research, under Contract DAAG29-77-C-0035.

**Research supported by US. National Heart and Lung Institute Contract NIH-NHLI-71-2243 from the National Institutes of Health, and by Public Health Research Grant 1 R01 CA 17107 from the National Cancer Institute.

1. INTRODUCTION

Of recent years there has been considerable interest in the use of multiplicative models for hazard rates, as a means whereby the influence of concomitant variables on survival can be expressed, and so allowed for in the analysis of survival data. Methods of analysis have been developed for estimation of parameters, reflecting dependence of survival on concomitant variables, which are robust with respect to the form of hazard rate function (provided the multiplicative model is appropriate). Use of these methods does involve sacrifice of some of the information typically available in survival data.

In this paper the nature and likely effects of the omitted information are studied, mainly by comparison of the "critical functions" of the parameters reflecting dependence on concomitant variables which have to be maximized to obtain maximum likelihood estimators, according as to what information is included and what further assumptions about the model are made.

2. NOTATION

We suppose that N individuals in all are under observation at some time or other during the study. We also suppose that observation continues over a single interval of time — that is, that no individual withdraws and then reenters the study later. The i-th individual will be denoted by (i); for each individual (i), the data include times of entry and of withdrawal or death (failure) and also values of s concomitant variables $(z_{1i}, \ldots, z_{si}) = \underset{\sim}{z}_i'$.

Time $t = 0$ may correspond to date of initiation of a treatment (as in many clinical studies) or to date of birth (as in some occupational studies). In the first type of study it will very often be the case that the time of entry is 0 for all individuals, but this need not be so in general.

If, among the N individuals, n are observed to fail during the course of the study, we denote the ordered failure times by $t_1 < t_2 < \cdots < t_n$. We denote the individual failing at time t_j by $(i(j))$. (Multiple failures, when $m_j (>1)$ failures are recorded at the same time, t_j, are discussed in Section 8.)

The set of individuals in the study at a time t — those "exposed to risk" of being observed to fail — is called the *risk set* at time t. In particular the risk set at ("just before") the j-th failure — at time t_j — is denoted by R_j. We also use the notation R_j' to denote the set of all individuals who are under observation for *at least part* of the interval $I_j \equiv (t_{j-1}, t_j]$ $(j = 1, \ldots, n)$.

3. FORMULATION OF THE PROBLEM

We are interested in estimating the survival distribution function (SDF)

$$S(t|\underset{\sim}{z}) = \Pr[T<t|\underset{\sim}{z}], \tag{1}$$

where T denotes failure time. Usually it is impracticable to have sufficient numbers of individuals with common (or even approximately common) sets of values of the z's, so it is necessary to make some assumptions about the way the z's might combine to affect the SDF. It might also appear necessary to make assumptions about the form of dependence on t, specifying at least a parametric model for this. While this is so, it is possible to estimate parameters reflecting dependence on the z's, without making detailed assumptions about the form of SDF as a function of t.

4. THE GENERAL MULTIPLICATIVE MODEL

These methods are based on use of the *general multiplicative model* for the hazard rate function

$$-d\log S(t|\underset{\sim}{z})/dt = \lambda(t|\underset{\sim}{z}) = \lambda(t)g(\underset{\sim}{z};\underset{\sim}{\beta}) , \qquad (2)$$

where $\lambda(t)$ is an "underlying" hazard rate of unspecified form (except that $\lambda(t) \geq 0$), while $g(\cdot)$ is a known function of $\underset{\sim}{z}$ and of r unknown parameters $\beta_1, \ldots, \beta_r$. Usually r is taken equal to s, and $g(\cdot)$ is assumed to be a function of $\sum_{u=1}^{s} \beta_u z_u$, but this is not essential. If $\underset{\sim}{z}$ varies with time, we can write $\underset{\sim}{z}(t)$ on the right hand side of (2).

5. MAXIMUM PARTIAL LIKELIHOOD ESTIMATORS

A commonly used likelihood function for estimating the β's utilizes information only on individuals in the risk sets R_j. Given that a failure does occur at t_j, among the risk set R_j, and supposing $z = z(t)$, the probability that the individual $(i(j))$ is the one which fails is

$$g(\underset{\sim}{z}_{i(j)}(t_j);\ \underset{\sim}{\beta})\Big[\sum_{\ell \in R_j} g(\underset{\sim}{z}_{\ell}(t_j);\ \underset{\sim}{\beta})\Big]^{-1} , \qquad (3)$$

where $\ell \in R_j$ means that individual (ℓ) belongs to R_j. This does not depend on $\lambda(t)$. Cox (1972) used the product

$$\prod_{j=1}^{n} \left[\frac{g(\underset{\sim}{z}_{i(j)}(t_j);\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R_j} g(\underset{\sim}{z}_\ell(t_j);\ \underset{\sim}{\beta})} \right] , \tag{4}$$

as a "partial likelihood" from which estimation of the β's can be found by maximizing this statistic. While this approach leads to estimators of the β's uninfluenced by $\lambda(t)$ it should be noted that:

(i) Only the information t_j, R_j and $(i(j))$ is used, that is, information on events inside the intervals I_j is ignored. (This is, why the term *"partial* likelihood" is applied to (4));

(ii) In order to estimate $S(t|\underset{\sim}{z})$ it is still necessary to introduce some assumptions about the form of $\lambda(t)$.

6. MAXIMUM LIKELIHOOD ESTIMATION

In order to approximate the form of $\lambda(t)$ it may be supposed that it remains *constant over each interval,* I_j, but can change from interval to interval. Under this assumption

$$\lambda(t) = \lambda_j \quad \text{for} \quad t_{j-1} < t \le t_j , \tag{5}$$

and the likelihood function for the data described in Section 1, with the model (2) is

$$L(\underset{\sim}{\lambda};\ \underset{\sim}{\beta}) = L(\lambda_1, \ldots, \lambda_n;\ \underset{\sim}{\beta}) = \prod_{j=1}^{n} L_j(\lambda_j;\ \underset{\sim}{\beta}) , \tag{6}$$

where

$$L_j(\lambda_j;\ \underset{\sim}{\beta}) = \lambda_j g(z_{i(j)}(t_j);\ \underset{\sim}{\beta}) \exp[-\lambda_j \sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_\ell(t);\ \underset{\sim}{\beta}) dt] , \tag{7}$$

and $I_{j\ell}$ denotes that part of I_j over which (ℓ) is observed. Note the difference between R_j' and $R_j - R_j'$ consists of R_j, *plus* all those individuals who withdraw.

We first maximize (6) with respect to the λ_j's (supposing the β's fixed) and then maximize the resultant value, $L(\hat{\underset{\sim}{\lambda}}(\underset{\sim}{\beta});\ \underset{\sim}{\beta})$, with respect to the β's.

Clearly

$$\max_{\lambda_1,\ldots,\lambda_n} L(\underset{\sim}{\lambda};\ \underset{\sim}{\beta}) = \prod_{j=1}^{n} [\max_{\lambda_j} L_j(\lambda_j;\ \underset{\sim}{\beta})] \ . \tag{8}$$

It is easy to show that

$$\max_{\lambda} \lambda e^{-\lambda w} = w^{-1} e^{-1} \ ,$$

corresponding to $\lambda = w^{-1}$. Hence $L(\underset{\sim}{\lambda};\ \underset{\sim}{\beta})$ is maximized by

$$\hat{\lambda}_j(\underset{\sim}{\beta}) = [\sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_{\ell}(t);\ \underset{\sim}{\beta})dt]^{-1} \ , \tag{9}$$

and the maximized value is

$$L(\hat{\lambda}(\beta);\ \beta) = e^{-n} \prod_{j=1}^{n} \left| \frac{g(\underset{\sim}{z}_{i(j)}(t_j);\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_{\ell}(t);\ \underset{\sim}{\beta})dt} \right. \ . \tag{10}$$

The maximum likelihood estimators of the β's are those values which maximize the critical function

$$e^{n}L(\hat{\underset{\sim}{\lambda}}(\underset{\sim}{\beta});\ \underset{\sim}{\beta}) = \prod_{j=1}^{n}\left[\frac{g(\underset{\sim}{z}_{i(j)}(t_j);\ \underset{\sim}{\beta})}{\sum_{\ell\epsilon R_j'}\int_{I_{j\ell}} g(\underset{\sim}{z}_{\ell}(t;\ \underset{\sim}{\beta})dt}\right], \tag{11}$$

Comparing (11) with (4) we see that the numerators are identical but the denominators differ,

$$\sum_{\ell\epsilon R_j} g(\underset{\sim}{z}_{\ell}(t_j);\ \underset{\sim}{\beta}) \quad \text{in (4)}$$

being replaced by

$$\sum_{\ell\epsilon R_j'}\int_{I_{j\ell}} g(\underset{\sim}{z}_{\ell}(t);\ \underset{\sim}{\beta})dt \quad \text{in (11).}$$

Whereas the quantity in (4) depends only on the risk set (R_j) and values of the concomitant variables "at t_j," in (11) it depends on the risk set (R_j') during I_j and the whole periods of exposure of individuals (ℓ) in I_j.

7. SOME IMPORTANT SPECIAL CASES

In practice $\underset{\sim}{z}(t)$ will not be known for all t. In many cases, only fixed values of concomitant variables are used — that is $\underset{\sim}{z}(t)$ does not depend on t, and so can be represented as $\underset{\sim}{z}$. If this is so,then

$$\int_{I_{j\ell}} g(\underset{\sim}{z}_{\ell};\ \underset{\sim}{\beta})dt = h_{j\ell}g(\underset{\sim}{z}_{\ell};\ \underset{\sim}{\beta}), \tag{12}$$

where $h_{j\ell}$ is the length of time ("person × time units exposed to risk") for which (ℓ) is under observation in I_j. For individuals under observation for the whole of I_j,

$$h_{j\ell} = t_j - t_{j-1} = h_j \quad . \tag{13}$$

If precise information on times of entry and withdrawal are not available, more or less arbitrary approximations must be used. If (ℓ) either enters *or* withdraws from the study (but does not do both) in I_j one might take $h_{j\ell} \doteq \frac{1}{2} h_j$; if (ℓ) both enters *and* withdraws in I_j, one might take $h_{j\ell} \doteq \frac{1}{3} h_j$. Sometimes there may be specific information on withdrawal and/or entry "habits" which might lead to modification of these formulae.

From (9) and (12) we see that

$$\hat{\lambda}_j(\underset{\sim}{\beta}) = \left[\sum_{\ell \varepsilon R'_j} h_{j\ell}\, g(\underset{\sim}{z}_\ell;\ \underset{\sim}{\beta}) \right]^{-1} \quad , \tag{14}$$

and the critical function (11) takes the form

$$\prod_{j=1}^{n} \left[\frac{g(\underset{\sim}{z}_{i(j)};\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R_j} h_{j\ell}\, g(\underset{\sim}{z}_\ell;\ \underset{\sim}{\beta})} \right] \quad . \tag{15}$$

This is a constant multiple $\left(\prod_{j=1}^{n} h_j^{-1} \right)$ of

$$\prod_{j=1}^{n} \left[\frac{g(\underset{\sim}{z}_{i(j)};\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R'_j} \theta_{j\ell}\, g(\underset{\sim}{z}_\ell;\ \underset{\sim}{\beta})} \right] \quad , \tag{16}$$

where $\theta_{j\ell} = h_{j\ell}/h_j$ is the proportion of I_j for which (ℓ) is under observation.

The critical function (partial likelihood) (4) is now

$$\prod_{j=1}^{n} \left[\frac{g(\underset{\sim}{z}_{i(j)};\ \underset{\sim}{\beta})}{\sum_{\ell \in R_j} g(\underset{\sim}{z}_{\ell};\ \underset{\sim}{\beta})} \right] . \tag{17}$$

Comparison of (16) and (17) shows that they differ in their denominators: — (a) the summation in (16) is over R_j', while that in (17, is over R_j, thus omitting individuals withdrawing in I_j; (b) the factors $\theta_{j\ell}$ in (16) reflect different proportions of I_j during which the corresponding individuals were under observations. To sum up, (17) could be obtained from (16) by (a) ignoring individuals who withdrew during I_j, and (b) supposing all individuals who enter (and do not withdraw) during I_j to have been under observation for the whole of the interval.

Situations when there are no new entries deserve, perhaps, special attention. These are typical in clinical trials, where t is the follow up time since initiation of the treatment.

For computational purposes, it might be useful to represent (16) and (17) in more convenient forms.

Let τ_ℓ denote the time of departure (by failure or withdrawal) for individual (ℓ).

Then

$$\theta_{j\ell} = \begin{cases} 0 & \text{for } \tau_\ell \le t_{j-1} \\ (\tau_\ell - t_{j-1})/h_j & \text{for } t_{j-1} < \tau_\ell < t_j \\ 1 & \text{for } \tau_\ell \ge t_j, \end{cases} \tag{18}$$

so that (16) takes the form

$$\prod_{j=1}^{n} \left[\frac{g(\underset{\sim}{z}_{i(j)};\ \underset{\sim}{\beta})}{\sum_{\ell=1}^{N} \theta_{j\ell}\, g(\underset{\sim}{z}_{\ell};\ \underset{\sim}{\beta})} \right] . \tag{19}$$

If we were to define

$$\theta_{j\ell} = \begin{cases} 0 & \text{if} \quad \tau > t_j \\ 1 & \text{if} \quad \tau \geq t_j \ , \end{cases} \tag{20}$$

formula (19) would correspond to (17).

8. MULTIPLE FAILURES

If several failures — m_j, say — are recorded at the same time, then, in practice, it means that the time unit of record is not sufficiently small to distinguish them. However, it is possible to establish a likelihood function, treating the failures as if they really did occur at the same time, and so to obtain maximum likelihood estimators.

We denote the m_j individuals failing at time t_j by $(i(j,1))$, $(i(j,2)), \ldots, (i(j,m_j))$ $(j = 1, 2, \ldots, n)$. Of course, we must have $m_1 + m_2 + \ldots + m_n \leq N$.

The likelihood function is

$$L(\lambda;\ \beta) = \prod_{n=1}^{n} \left[\lambda_j^{m_j} \left\{ \prod_{k=1}^{m_j} g(\underset{\sim}{z}_{i(j,k)}(t_j);\ \underset{\sim}{\beta}) \right\} \exp\left\{ -\lambda_j \sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_{\ell}(t);\ \underset{\sim}{\beta})\, dt \right] \right\} . \tag{21}$$

Maximizing with respect to $\underset{\sim}{\lambda}$ gives

$$\hat{\lambda}_j(\underset{\sim}{\beta}) = m_j \left[\sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_\ell(t);\ \underset{\sim}{\beta})\,dt \right]^{-1}, \tag{22}$$

and the maximized value of the likelihood is

$$L(\hat{\lambda}(\underset{\sim}{\beta});\ \underset{\sim}{\beta}) = e^{-n} \left(\prod_{j=1}^{n} m_j^{m_j} \right) \prod_{j=1}^{n} \left[\frac{\prod_{k=1}^{m_j} g(\underset{\sim}{z}_{i(j,k)}(t_j);\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_\ell(t);\ \underset{\sim}{\beta})\,dt} \right]. \tag{23}$$

If $I_{j\ell} \equiv I_j$ in (22), we obtain Breslow's (1972) formula. The β's are obtained by maximizing the critical function

$$e^{n} \left(\prod_{j=1}^{n} m_j^{m_j} \right)^{-1} L(\underset{\sim}{\hat{\lambda}}(\underset{\sim}{\beta});\ \underset{\sim}{\beta}) = \prod_{j=1}^{n} \left[\frac{\prod_{k=1}^{m_j} g(\underset{\sim}{z}_{i(j,k)}(t_j);\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R_j'} \int_{I_{j\ell}} g(\underset{\sim}{z}_\ell(t);\ \underset{\sim}{\beta})\,dt} \right], \tag{24}$$

which differs from (11) only by substitution of

$$\prod_{k=1}^{m_j} g(\underset{\sim}{z}_{i(j,k)}(t_j);\ \underset{\sim}{\beta}) \quad \text{for} \quad g(\underset{\sim}{z}_{i(j)}(t_j);\ \underset{\sim}{\beta}),$$

in the numerator.

If the z's do not depend on t, then in place of (24) we can use the critical function

$$\prod_{j=1}^{n} \left[\frac{\prod_{k=1}^{m_j} g(\underset{\sim}{z}_{i(j,k)};\ \underset{\sim}{\beta})}{\sum_{\ell \varepsilon R_j'} \theta_{j\ell}\, g(\underset{\sim}{z}_\ell;\ \underset{\sim}{\beta})} \right]. \tag{25}$$

9. ESTIMATION OF SURVIVAL FUNCTION WHEN THERE ARE NO COVARIATES

When the covariates are absent so that $g(\cdot) \equiv 1$, formula (22) gives

$$\hat{\lambda}_j = m_j \left\{ \sum_{\ell \in R'_j} h_{j\ell} \right\}^{-1} . \tag{26}$$

The corresponding estimate of the SDF (for $t_{j-1} \le t < t_j$) is

$$\hat{S}(t) = \exp\left\{ -\sum_{f=1}^{j-1} \hat{\lambda}_f h_f - \hat{\lambda}_j (t - t_{j-1}) \right\}$$

$$= \exp\left[-\sum_{f=1}^{j-1} m_f \left(\sum_{\ell \in R'_f} \theta_{f\ell} \right)^{-1} - m_j (t - t_{j-1}) h_j^{-1} \left(\sum_{\ell \in R'_j} \theta_{j\ell} \right) \right] \tag{27}$$

for $t_{j-1} \le t < t_j$ $(j = 1, \ldots, n)$.

For $t > t_n$ we have $\hat{S}(t) = \hat{S}(t_n)$ formally, but there may be relatively few data for these t (they correspond to individuals under observation for longer periods than that to the last failure).

If $\theta_{f\ell} \equiv 1$ in (16) — that is, there are no new entries, and if $R'_f \equiv R_f$ — that is, there are no withdrawals — then (27) gives

$$\hat{S}(t) = \exp\left[-\sum_{f=1}^{j-1} m_f R_f^{-1} - m_j R_j^{-1} (t - t_{j-1}) h_j^{-1} \right], \quad t_{j-1} \le t < t_j . \tag{28}$$

By comparison, the Kaplan-Meier estimator of $S(t)$ is

$$\hat{S}^*(t) = \prod_{f=1}^{j-1} (1 - mR_f^{-1}) \qquad (t_{j-1} \le t < t_j), \tag{29}$$

where R_f is the number of individuals in R_f.

Another approximation, given by Thompson (1977), leads to a formula like (29), but with R_f increased by one-half of the number of withdrawals in I_f. Of course, when there are no multiple failures, formulae (22) through (29) are valid with m_f's replaced by 1.

10. EXAMPLE

The data in Table 1 are adopted from Crowley and Hsu (1974) ("Covariance Analysis of Heart Transplant Survival Data," Technical Report No. 2, Division of Biostatistics, Stanford University). These are survival data for $N = 64$ patients who had heart transplants. The date of entry corresponds here to the date of operation, so that the time at entry is equal to zero for each patient in the follow-up. The investigation covers the period January 6, 1968 to April 1, 1974 , that is 2276 days. The time τ_ℓ (in days) is the time beyond which individual (ℓ) was no longer under observation (after death or withdrawal). The data are arranged in increasing order of τ_ℓ (column 3 of Table 1). Note that for patient (1) $\tau_1 = 0^+$; we have assigned arbitrarily the value 0.5 days. The numbers in parentheses in column 1 are the ID numbers of the patients, assigned to them at entry. Column 2 gives the values of the indicator variable δ_ℓ, equal to 1 if individual (ℓ) is observed to die, and 0 otherwise. There were $n = 39$ deaths, at times t_j. (Note that when $\delta_\ell = 1$, $\tau_{\ell(j)} = t_j$.) The values of τ_ℓ, when $\delta_\ell = 0$ are distinguished by asterisk.

Out of 8 concomitant variables given in the original paper, only two are shown here: age at entry, z_1 (in years), and the so called "mismatch score", z_2, based on tissue and blood typing of each patient.

We use these data for illustrating the methods of estimation. There is no claim to evaluate the study as a whole.

The model used here is simply

$$\lambda(\underset{\sim}{t}_j;\ \underset{\sim}{z}_j) = \lambda_j \exp(\beta_1 z_{1j} + \beta_2 z_{2j}).$$

The parameters β_1, β_2, and the λ_j's are estimated by two approaches.

(a) Using the information on withdrawal times, with the risk sets R_j'. The likelihood function (19), with $\theta_{j\ell}$ defined in (18), is used. We notice a multiplicity at t_{17} $(m_{17} = 2)$. The estimates of β's and their estimated standard errors are:

$$\hat{\beta}_1 = 0.0705, \quad S.D.(\hat{\beta}_1) = 0.0233;$$

$$\hat{\beta}_2 = 0.6644, \quad S.D.(\hat{\beta}_2) = 0.2876 .$$

(b) Ignoring the information on withdrawal time, that is, using only the risk sets, R_j. The likelihood function is still given by (19), but with $\theta_{j\ell}$ defined by (20). The results are:

$$\hat{\beta}_1 = 0.0699, \quad S.D.(\hat{\beta}_1) = 0.0234$$

$$\hat{\beta}_2 = 0.6817, \quad S.D.(\hat{\beta}_2) = 0.2883 .$$

We also evaluate the estimated survival functions ignoring covariates, that is when $g(\cdot) = 1$, in three different ways.

(a) $\widehat{S1}(t)$ - utilizing information on withdrawal times with sets R_j';

(b) $\widehat{S2}(t)$ - assuming that the withdrawal time is at the beginning of the interval for all those who withdraw in the interval, with sets R_j.

Since there were no new entries, $\lambda_j(t)$, is estimated from the formula

$$\hat{\lambda}_j(t) = m_j [\sum_{\ell=1}^{64} \theta_{j\ell} h_j]^{-1} , \qquad (30)$$

with $\theta_{j\ell}$ defined by (18) for appraoch (a), and by (20) for approach (b). Then $\widehat{S1}(t)$ and $\widehat{S2}(t)$ are calculated from (27) and (28), respectively.

(c) $\widehat{S^*}(t)$ - the product-limit estimates, with sets R_j.

The resulting SDF's are given in the three last columns of Table 1. There are not many withdrawals between any two deaths (this is especially true for early deaths). In these circumstances is to be expected that these three estimates of SDF do not differ greatly as is, indeed, the case.

ACKNOWLEDGEMENT

We would like to thank Paul Wan for computation in the Example.

REFERENCES

Breslow, N.E. (1972). Discussion of D.R. Cox (1972) *J.R. Statist. Soc. Ser. B* **33**, 216-217.

Cox, D.R. (1972). Regression models and life tables. *J.R. Statist. Soc. Ser. B* **33**, 187-202.

Thompson, W.A. (1977). On the treatment of grouped observations in life studies. *Biometrics* **33**, 463-470.

TABLE 1

Estimation of Survival Function for Heart Transplant Data

ℓ	(ID)	δ_ℓ	τ_ℓ	j	t_j	h_j	$z_{1\ell}$	$z_{2\ell}$	R'_j	R_j	$\widehat{S1}(\tau_\ell)$	$\widehat{S2}(\tau_\ell)$	$\widehat{S^*}(\tau_\ell)$
1	(38)	1	0.5	1	0.5	0.5	41.5	0.87	64	64	.9844	.9844	.9844
2	(28)	1	1	2	1	0.5	54.3	0.47	63	63	.9690	.9690	.9688
3	(100)	0	1*				35.2	0.67			.9690	.9690	↓
4	(4)	1	3	3	3	2	40.4	1.66	62	61	.9532	.9532	.9531
5	(20)	1	10	4	10	7	45.3	2.76	60	60	.9375	.9375	.9372
6	(74)	1	12	5	12	2	29.2	0.61	59	59	.9217	.9217	.9214
7	(98)	0	13*				28.9	0.77			.9164	.9163	↓
8	(3)	1	15	6	15	3	54.3	1.11	58	57	.9058	.9057	.9052
9	(18)	1	23	7	23	8	46.9	2.05	56	56	.8898	.8897	.8890
10	(70)	1	25	8	25	2	53.0	1.68	55	55	.8737	.8736	.8728
11	(90)	1	26	9	26	1	52.5	0.82	54	54	.8577	.8576	.8567
12	(79)	1	29	10	29	3	54.0	1.08	53	53	.8417	.8416	.8405
13	(92)	0	30*				45.8	0.16			.8400	.8399	↓
14	(22)	1	39	11	39	10	42.8	1.38	52	51	.8254	.8252	.8241
15	(45)	1	44	12	44	5	36.3	0.00	50	50	.8090	.8089	.8076
16	(10)	1	46	13	46	2	42.5	0.61	49	49	.7927	.7926	.7911
17	(37)	1	47	14	47	1	61.7	0.87	48	48	.7763	.7762	.7746
18	(83)	1	48	15	48	1	53.3	3.05	47	47	.7600	.7599	.7581
19	(87)	1	50	16	50	2	46.4	2.25	46	46	.7436	.7435	.7416
20	(47)	1	51}	17	51}	1	47.2	1.38	45}	45}	.7113}	.7112}	.7087}
21	(55)	1	51				52.5	1.51					
22	(36)	1	54	18	54	3	49.1	2.09	43	43	.6950	.6949	.6922
23	(32)	1	60	19	60	6	64.4	0.69	42	43	.6786	.6785	.6757
24	(73)	1	63	20	63	3	56.4	2.16	41	41	.6623	.6622	.6592
25	(13)	1	64	21	64	1	54.6	1.89	40	40	.6459	.6458	.6428
26	(68)	1	65	22	65	1	45.3	1.68	39	39	.6296	.6295	.6293
27	(65)	1	66	23	66	1	51.3	1.12	38	38	.6132	.6131	.6098
28	(89)	1	68	4	68	2	51.4	1.33	37	37	.5969	.5968	.5933
29	(97)	0	109*				23.6	1.78			.5853	.5850	↓
30	(11)	1	127	25	127	59	48.0	0.36	36	35	.5804	.5800	.5764
31	(24)	1	136	26	136	9	52.0	1.62	34	34	.5635	.5632	.5594
32	(94)	1	161	27	161	25	43.9	1.20	33	33	.5467	.5463	.5425
33	(96)	0	166*				26.3	0.46			.5454	.5450	↓
34	(67)	1	228	28	228	67	19.8	1.02	32	31	.5294	.5290	.5250
35	(93)	0	236*				47.8	0.33			.5237	.5232	↓
36	(51)	1	253	29	259	25	48.8	1.08	30	29	.5117	.5111	.5069
37	(16)	1	280	30	280	17	49.5	1.12	28	28	.4937	.4931	.4888
38	(84)	1	297	31	297	17	42.8	0.60	27	27	.4758	.4752	.4707
39	(78)	0	304*				49.3	0.81			.4705	.4699	↓
40	(58)	1	322	32	322	25	48.1	1.82	26	25	.4573	.4566	.4518
41	(88)	0	338*				55.4	0.68			.4558	.4549	↓
42	(86)	0	388*				48.9	1.44			.4511	.4497	
43	(81)	0	438*				52.9	1.94			.4465	.4446	
44	(80)	0	455*				46.5	1.41			.4450	.4428	
45	(76)	0	498*				52.2	1.70			.4410	.4385	
46	(64)	1	551	33	551	229	48.9	0.12	24	19	.4362	.4332	.4280
47	(71)	0	588*				46.5	0.97			.4235	.4197	↓
48	(72)	0	591*				26.7	1.46			.4225	.4186	
49	(7)	1	624	34	624	73	51.0	1.32	18	16	.4114	.4069	.4013
50	(69)	0	659*				48.0	1.20			.4020	.3974	↓
51	(23)	1	730	34	730	106	58.4	0.96	15	14	.3837	.3789	.3726
52	(63)	0	814*				32.7	1.93			.3606	.3547	↓
53	(30)	1	836	36	836	106	45.0	1.58	13	12	.3548	.3486	.3416
54	(59)	0	837*				41.6	0.19			.3546	.3483	↓
55	(56)	0	874*				38.1	0.98			.3457	.3394	
56	(46)	1	994	37	994	158	48.5	0.81	11	9	.3184	.3119	.3036
57	(21)	1	1024	38	1024	30	43.4	1.13	8	8	.2810	.2753	.2657
58	(49)	0	1105*				36.8	1.35			.2696	.2619	↓
59	(41)	0	1263*				45.5	0.98			.2486	.2377	
60	(14)	1	1350	39	1350	326	54.1	0.87	7	5	.2378	.2254	.2205
61	(40)	0	1366*				48.6	0.75					↓
62	(33)	0	1536*				49.0	0.91					
63	(34)	0	1548*				40.5	0.38					
64	(25)	0	1774*				33.2	1.06					

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Estimation in the General Multiplicative Model for Survival	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1187
7. AUTHOR(s) Norman L. Johnson and Regina C. Elandt-Johnson	8. CONTRACT OR GRANT NUMBER(s) DAAG29-77-C-0035
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AROD, Research Triangle Park, NC	12. REPORT DATE July 1978
	13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Estimation, Multiplicative mortality model, Maximum likelihood, Partial likelihood, Ties, Multiple deaths

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The general multiplicative model (formula (2)) represents the hazard function as the product of an "underlying hazard rate, $\lambda(t)$, of unspecified form and a certain function of known form, $g(z; \beta)$, where z is a vector of concomitant variables, and β is a vector of unknown parameters. Assuming that $\lambda(t)$ can be approximated by a constant between any two consecutive failures, the general forms of likelihood function are derived (formulae (6) and (7), or (9) and (11)). The likelihood utilizes the available information on the time of exposure to risk of each individual (until failure or withdrawal). Special cases, when the z's do

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. not depend on t are discussed in some detail (Section 7). Multiple failures are handled in a simple manner - no ordering of failures is required (Section 8). Estimation of empirical survival function when there are no covariates is discussed in Section 9. An example using heart transplant data, is given (for illustrative purpose only).