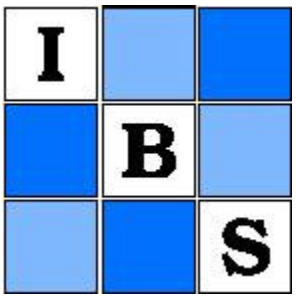

BIOMETRICS 37, 67–78
March 1981

Evaluating Serial Cancer Marker Studies in Patients at Risk of Recurrent Disease

Mitchell H. Gail

National Cancer Institute, Landow Building, 7910 Woodmont Avenue, Bethesda, Maryland 20205, U.S.A.

SUMMARY

The proportional hazards model of Cox (1972, *Journal of the Royal Statistical Society, Series B* **34,** 187–220), with a time-dependent covariate, is used to analyze serial cancer marker data. A particular advantage of this method is the ease with which missing marker data are handled. Analysis of a real data set shows that high levels of the cancer marker carcinoembryonic antigen (CEA) are associated with increased risk of death in patients with resected colorectal cancer. Several aspects of CEA marker history are analyzed, including CEA level at death time t, CEA level 200 days prior to time t, and whether or not CEA exceeded 5 ng/ml prior to t. Methods to test the hypothesis of no marker effect and to give estimates and confidence intervals for model parameters are outlined both for continuous and for grouped time-to-response data. For grouped data a likelihood ratio test of the proportional hazards assumption is suggested.

1. Introduction

A variety of new medical tests, known as cancer markers, are being used to monitor the course of disease in patients under treatment. One of these, the beta subunit of human chorionic gonadotropin discussed by Pastorfide, Goldstein and Kosasa (1974), is thought to be an excellent guide to therapy. Alpha-fetoprotein may have some value in monitoring hepatic and testicular cancer (see Waldmann and McIntire, 1974), and the value of serial carcinoembryonic antigen (CEA) mmsurements in the early diagnosis of recurrent colorectal cancer is a matter of continued study and debate, as evidenced by the various results of Moertel, Schutt and Go (1978), Mach *et al.* (1974) and Booth *et al.* (1974). Especially for markers such as CEA which exhibit a variable relationship between marker value and extent of disease, there is a need for statistical methods to determine whether risk of recurrence or death is indeed related to the monitoring test. For this purpose we use the proportional hazards model with time-dependent covariates which has been defined and developed by Cox (1972) and Crowley and Hu (1977), and used implicitly by Mantel and Byar (1974). Applying these methods to serial CEA measurements in patients with resected colorectal cancer, we find strong evidence that increased risk of death is associated with higher CEA levels.

Our strategy is to define various functionals $\mathbf{Z}(t)$ which depend on the marker history up to time t. Regarding these functionals as time-dependent covariates, we test whether they are related to risk of death (or recurrence). Suppose γ is the value of the marker at time t. Then the marker history for an individual is $M=\{\gamma(\tau):0\leqslant\tau\leqslant t\}$. The physician can focus on particular aspects of serial marker behavior by defining particular vector valued functionals $\mathbf{Z}(t)$ on the set $\mathcal{M}$ of possible marker histories M. Some examples in

Key words: Cancer marker; Monitoring test; Proportional hazards model; Time-dependent covariates; Carcinoembryonic antigen; Colorectal cancer.

which $\mathbf{Z}(t)$ is a scalar are:

$$Z_1(t) = \gamma(t),$$
$$Z_2(t) = \gamma(t-\omega),$$
$$Z_3(t) = \begin{cases} 1 & \text{if } \sup \gamma(t) \geqslant \eta, \quad 0 \leqslant \tau \leqslant t, \\ 0 & \text{otherwise,} \end{cases}$$
$$Z_4(t) = \{\gamma(t) - \gamma(t-\Delta)\}/\Delta.$$

Here $Z_1(t)$ is simply the marker value at t, and $Z_1(t)$ may be used to study the question: 'Are those with elevated marker values at time t at higher risk at that time?' Likewise $Z_2(t)$ may be used to study the question: 'Are those with elevated marker values at time $t-\omega$ at higher risk at time t?' The covariate $Z_3(t)$ might be used to see whether any previous marker elevation puts one at higher risk, and the covariate $Z_4(t)$ might be used to see if a high rate of increase in the marker has grave prognostic significance. These examples illustrate the ability to focus on any aspect of the marker history.

It is often convenient to use discrete versions of the functionals above. In the example discussed in §3, we consider four ranges of CEA values, $[0, 2.5]$, $(2.5, 5]$, $(5, 10]$ and $(10, \infty)$. (The unit of measurement for CEA is ng/ml throughout this paper.) We define the marker state vector $\mathbf{Z}'(t) = \{z_2(t), z_3(t), z_4(t)\}$ such that the four CEA ranges above correspond respectively to discrete states $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. These four states are not necessarily shown in order of increasing hazard, and this formulation allows one to study the relative risks comparing any two such states. We also test for trend in §3 by assigning scores 1, 2, 3 or 4 to the scalar functional $Z_1(t)$ according as CEA is in the ranges $[0, 2.5]$, $(2.5, 5]$, $(5, 10]$ or $(10, \infty)$. Other discrete versions of $Z_1(t)$ with different scores could be used.

The methods given in §2 were chosen to accommodate the following generic features of the serial marker data problem:

(i) The risk of death may be influenced by other prognostic factors which could obscure the effect of the serial marker.
(ii) There is often insufficient information to justify a particular parametric model for the analysis.
(iii) The time to death data is variably censored on the right.
(iv) The marker value $\gamma(t)$ is only measured at a finite number of points.
(v) Occasionally one has no idea what value to assign $\gamma(t)$ or $\mathbf{Z}(t)$ because no proximate values are available.

The first three problems arise whenever one attempts a covariate analysis on survival data, and the semiparametric approach of Cox (1972) is well adapted to these problems. We adjust for other prognostic factors by stratification and allow a separate nuisance hazard function for each stratum. Problem (iv) requires the data analyst to define an interpolation convention to assign values $\gamma(t)$ for times t intermediate between observations. The partial likelihood method of Cox (1972, 1975) is particularly useful for Problem (v), because it allows patients to contribute to 'risk sets' when, and only when, a valid marker measurement is available. To make this concept precise, the analyst must define 'proximate.' A marker measurement $\gamma(t)$ is said to be 'censored' if no proximate measurements are available; otherwise, $\gamma(t)$ is said to be a 'valid marker measurement.' Likewise, the experiment yields a 'valid functional measurement' or a 'censored functional measurement' according as sufficient proximate observations are or are not available to determine $\mathbf{Z}(t)$. For the same set of marker measurement, some functionals, such as the slope $Z_4(t)$, may be censored and others, such as $Z_1(t)$ may be valid. We now introduce some related

terms which are needed to discuss the information obtained from a marker study. A death (or event) is said to be 'valid' if the patient who dies at time t has a valid functional measurement $\mathbf{Z}(t)$. A death is said to be 'informative' if it is valid and if, at time t of the death, at least two patients are at risk with different valid functional measurements $\mathbf{Z}(t)$.

We assume that the timing of marker measurements and the procedures used for making marker measurements are not related to patient status, Otherwise, censorship of $\mathbf{Z}(t)$ might be informative of patient status, which would invalidate the methods in §2. Likewise, we assume an independent censorship mechanism for the time-to-death data.

These considerations allow us to define the risk set at the time t of an informative death as the set of patients who are alive and under observation at time t and who have a valid functional measurement at time t. A patient may fail to be at risk at time t_1 and later be at risk at time $t_2 > t_1$ if additional marker data become available. Thus, the partial likelihood used in §2 not only ignores the time-censorship mechanism as in Cox (1972, 1975) but also omits terms relating to the questions when or if a valid marker measurement is made.

In reporting such analyses, it is important that the functionals $\mathbf{Z}(t)$ and related conventions for interpolation and for determination of valid measurements be explicitly defined.

2. Methods

2.1 *Sparse Continuous Data*

The hazard at time t for an individual from stratum i with covariate $\mathbf{Z}(t)$ is assumed to be

$$h(t) = h_{0i}(t) \exp\{\boldsymbol{\alpha}'\mathbf{Z}(t)\}, \tag{1}$$

where $h_{0i}(t)$ is a nuisance hazard function characteristic of stratum i. Here $\boldsymbol{\alpha}' = (\alpha_2, \alpha_3, \ldots, \alpha_k)$ and $\mathbf{Z}(t)$ are so defined that $\boldsymbol{\alpha}'\mathbf{Z}(t) = 0$ for covariate state 1 and $\boldsymbol{\alpha}'\mathbf{Z}(t) = \alpha_k$ for covariate state k, $k = 2, 3, \ldots, K$. As mentioned earlier, this formulation permits study of the case where the marker can be in one of K states, not necessarily in order of increasing hazard. Note that the time-dependent covariate effect is assumed constant over strata and through time for each stratum.

Let N_{ijk} be the number at risk (alive, under observation, and having a valid functional measurement) in covariate state k at the time of informative death(s) j in stratum i for $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J_i$, $k = 1, 2, \ldots, K$, and let D_{ijk} be the number who died at that time. From Cox (1972) and Breslow (1975), the log partial likelihood is approximated by

$$\sum_{i=1}^{I} \sum_{j=1}^{J_i} \sum_{k=1}^{K} \left\{ \alpha_k D_{ijk} - D_{ij+} \ln \sum_{l=1}^{K} N_{ijl} \exp(\alpha_l) \right\}, \tag{2}$$

where $\alpha_1 \equiv 0$. Routine likelihood calculations lead to estimates $\hat{\boldsymbol{\alpha}}$, and the $K-1$ degree of freedom likelihood ratio statistic may be used to test the null hypothesis of no marker effect, $H_0: \boldsymbol{\alpha} = \mathbf{0}$. By performing the analysis separately for each stratum and allowing a parameter vector $\boldsymbol{\alpha}_i$ for each stratum, one can obtain a likelihood ratio statistic with $(K-1)(I-1)$ degrees of freedom for testing homogeneity of covariate effect over strata.

Mantel and Byar (1974) use an alternative statistic to test $H_0: \boldsymbol{\alpha} = \mathbf{0}$. Their statistic can be recognized as a score statistic from (2), as in Cox (1972). To define this statistic, note that the null expectation (conditional on the observed numbers at risk) of component k of the vector $\mathbf{X}_{ij}' = (D_{ij2}, D_{ij3}, \ldots, D_{ijK})$ is given by

$$\mathrm{E}(D_{ijk}) = D_{ij+} N_{ijk} / N_{ij+} \equiv D_{ij+} \rho_{ijk}. \tag{3}$$

The covariance of $\mathbf{X}_{ij}$ is the $(K-1)\times(K-1)$ matrix $\mathbf{V}_{ij}$ with diagonal elements

$$D_{ij+}\rho_{ijk}(1-\rho_{ijk})(N_{ij+}-D_{ij+})/(N_{ij+}-1)$$

and off-diagonal elements

$$-D_{ij+}\rho_{ijk}\rho_{ijl}(N_{ij+}-D_{ij+})/(N_{ij+}-1) \quad \text{for} \quad k \neq l.$$

Letting $\mathbf{V}=\sum_{i,j}\mathbf{V}_{ij}$ and $\mathbf{Y}=\sum_{i,j}\{\mathbf{X}_{ij}-\mathrm{E}(\mathbf{X}_{ij})\}$, the Mantel–Haenszel test statistic for H_0 is

$$MH=\mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}, \tag{4}$$

a special case of which is given in Mantel and Haenszel (1959).

To test for trends, suppose instead that

$$h(t)=h_{0i}(t)\exp\{\beta Z(t)\}, \tag{5}$$

where now $Z(t)$ is a scalar. The scalar $Z(t)$ might be continuous, or it might be assigned ordered discrete values $z_1<z_2<z_k$ such as $1, 2, 3, \ldots, K$. Again, likelihood methods follow from the log partial likelihood

$$\sum_{i=1}^{I}\sum_{j=1}^{J_i}\left\{\sum_{k=1}^{K}\beta z_k D_{ijk}-D_{ij+}\ln\sum_{l=1}^{K}N_{ijl}\exp(\beta_{zl})\right\}. \tag{6}$$

A score test of trend may be based on the statistic

$$T=\sum_{i,j,k}z_k(D_{ijk}-D_{ij+}\rho_{ijk}), \tag{7}$$

as in Mantel (1963) and Tarone (1975). Under $H_0'\colon\beta=0$, T has mean zero and variance

$$\mathrm{var}\,(T)=\sum_{i,j}D_{ij+}(N_{ij+}-D_{ij+})(N_{ij+}-1)^{-1}\left\{\sum_k z_k^2\rho_{ijk}-\left(\sum_k z_k\rho_{ijk}\right)^2\right\}. \tag{8}$$

The statistic $T^2/\mathrm{var}\,(T)$, which has an asymptotic chi square distribution with one degree of freedom under H_0', yields a two-tailed test.

2.2 *Grouped Data*

If the data are grouped into actuarial intervals containing sizable numbers of events, one must modify the methods used for sparse data. Grouped data do afford an opportunity to test the proportional hazards models (1) and (5). Following Prentice and Gloeckler (1978), we compute the probability of dying in actuarial interval j of stratum i, namely $[\tau_{j-1}^i, \tau_j^i)$, from

$$\begin{aligned}Q_{ijk}&=1-\exp\left\{-\int_{\tau_{j-1}^i}^{\tau_j^i}h_{0i}(t)\exp(\alpha_k)\,dt\right\}\\&\equiv 1-P_{ij}\ **\ \exp(\alpha_k),\end{aligned}$$

where the double asterisk indicates exponentiation. Letting N_{ijk}, D_{ijk} and $S_{ijk}=N_{ijk}-D_{ijk}$ represent the numbers at risk at τ_{j-1}^i, dying in the interval, and surviving the interval, respectively, we obtain the partial likelihood

$$\mathscr{L}(\omega)=\prod_{i=1}^{I}\prod_{j=1}^{J_i}\prod_{k=1}^{K}\binom{N_{ijk}}{D_{ijk}}\{1-P_{ij}\ **\ \exp(\alpha_k)\}^{D_{ijk}}P_{ij}^{S_{ijk}\exp(\alpha_k)}, \tag{10}$$

which may be used to obtain estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{P}_{ij}$. Several remarks follow.

Inference based on (19) could conceivably break down if each interval contained only a few deaths because the number of nuisance parameters P_{ij} could proliferate too rapidly.

If W_{ijk} withdrawals occur within an interval, one can replace N_{ijk} in (10) by the 'effective' number at risk $N_{ijk} - \frac{1}{2}W_{ijk}$. This is a reasonable procedure if the death and censorship times are independent as in Gail (1975).

We assume without loss of generality that only those L intervals satisfying

$$0 < \sum_k D_{ijk} < \sum_k N_{ijk} \tag{11}$$

and

$$N_{ijk} > 0 \text{ and } N_{ijl} > 0 \text{ for some } k \text{ and some } l \neq k \tag{12}$$

are included in the product (10). Only such intervals give information about the parameters $\boldsymbol{\alpha}$ and are therefore termed 'informative intervals'. Equation (12) requires that valid deaths be informative, and equation (11) requires that only those intervals containing at least one informative death and one survivor be included in the likelihood (10). There are therefore $L+(K-1)$ parameters in the model (10) corresponding to P_{ij} and $\boldsymbol{\alpha}$ respectively. To test the model one can consider the alternative unconstrained likelihood

$$\mathscr{L}(\Omega) = \prod_{i=1}^{I} \prod_{j=1}^{J_i} \prod_{k=1}^{K} \binom{N_{ijk}}{D_{ijk}} (1-P_{ijk})^{D_{ijk}} P_{ijk}^{S_{ijk}} \tag{13}$$

based on KL degrees of freedom. The resulting likelihood ratio statistic which is asymptotically distributed as a chi square variate with $KL-K-L+1=(K-1)(L-1)$ degrees of freedom, is

$$SS_1 = 2\sum_i^I \sum_j^{J_i} \sum_k^K (D_{ijk} \ln [\hat{Q}_{ijk}/\{1-\hat{P}_{ij} * * \exp(\hat{\alpha}_k)\}] + S_{ijk} \ln \{\hat{P}_{ijk}/\hat{P}_{ij} * * \exp(\hat{\alpha}_k)\}), \tag{14}$$

where $\hat{Q}_{ijk} = D_{ijk}/N_{ijk}$.

If one applies these methods separately to each stratum, one obtains I separate statistics SS_{1i}, each with (J_i-1) $(K-1)$ degrees of freedom, to test the proportional hazards model for Stratum i. The statistic SS_1, in (14), tests not only that the proportional hazards model is valid within each stratum but also that the marker effect is homogeneous across strata. The quantity $SS_1 - \sum_i SS_{1i}$, with $I-1$ degrees of freedom, may be used to test homogeneity of $\boldsymbol{\alpha}$ across strata.

Methods for solving equation (10) and estimating the covariance $\boldsymbol{\Sigma}_{11}$ of $\hat{\boldsymbol{\alpha}}$ are given in the Appendix.

Grouped data may also be used to test the trend model (5) and to estimate β and test $\beta=0$ in that model. Such analyses are based on the partial likelihood (10) with βz_k in place of α_k. Details are in the Appendix.

The score statistics given above are valid for grouped data, and, indeed, approximate the score statistics derived from (10).

3. A Worked Example to Evaluate Carcinoembryonic Antigen as a Cancer Marker

Drs T. M. Chu, E. D. Holyoke and P. Lavin kindly provided data on patients with resected colorectal cancer. The endpoint was time to death, and the purpose of the investigation was to see whether elevations in CEA, following cancer removal, are associated with increased risk of death. For each patient the relevant data consisted of the

initial Dukes grade of disease and preoperative CEA value, which are important stratifying variables, and a sequence of follow-up dates, each of which is associated with an indication of whether the patient survived to that date and, except for missing values, a CEA measurement. Four strata were defined, based on the previous results of Chu *et al.* (1976). Stratum 1 contained 26 patients with initial Dukes Grade A or B and initial CEA <2.5 ng/ml. Stratum 2 contained 11 patients with initial Dukes Grades A or B and CEA $\geqslant 2.5$ ng/ml. Stratum 3 contained 24 patients with Dukes Grade C and CEA $<$ 2.5 ng/ml, and Stratum 4 contained 13 patients with Dukes Grade C and CEA $\geqslant$ 2.5 ng/ml. There were nine deaths in Stratum 1, five deaths in Stratum 2, 11 deaths in Stratum 3 and nine deaths in Stratum 4.

The first functional we studied was the vector $\mathbf{Z}'(t)=\{z_2(t), z_3(t), z_4(t)\}$ as in §1. The object was to determine the relative risk, for example, of a patient in state (0, 0, 1) at time t (i.e. CEA >10 ng/ml) compared to that of an otherwise similar patient in state (0, 0, 0) (i.e. CEA $\leqslant 2.5$ ng/ml). The CEA value at t, $\gamma(t)$, was taken to be the nearest prior CEA value, provided that this prior measurement occurred no more than 100 days before t. Otherwise $\gamma(t)$ was taken as 'censored', which renders that patient ineligible for the risk set at t. These conventions define a time 'window' for valid measurements and a step function interpolation to extend $\gamma(\cdot)$ to values of t between observations. The values of N_{ijk} and D_{ijk} at the time of each informative death are shown in Table 1. States (0, 0, 0), (1, 0, 0), (0, 1, 0) and (0, 0, 1) correspond respectively to $k=1, 2, 3$ and 4. Only the 17 deaths in Table 1 were informative. Of the 34 deaths, 14 were censored, three were valid but uninformative, and 17 were informative. Each time 'interval' contained only one death, because the data were not grouped, and each of the intervals indicated satisfy (11) and (12). It appears that deaths in Strata 3 and 4 occur most often from states $k=3$ or 4 which correspond to CEA values in the intervals $(5, 10]$ or $(10, \infty)$.

Table 1

The quantities $(D_{ij1}/N_{ij1}, D_{ij2}/N_{ij2}, D_{ij3}/N_{ij3}, D_{ij4}/N_{ij4})$ *for carcinoembryonic antigen marker data for death following resection of colorectal cancer*

Stratum 1		
(1/20, 0/2, 0/0, 0/0)	(1/15, 0/0, 0/2, 0/2)	(0/3, 0/0, 1/1, 0/0)
Stratum 2		
(0/5, 0/1, 0/0, 1/1)		
Stratum 3		
(1/17, 0/2, 0/2, 0/2)	(0/15, 1/2, 0/1, 0/1)	(0/10, 0/1, 1/2, 0/1)
(0/8, 0/0, 0/1, 1/4)	(0/7, 0/0, 0/1, 1/3)	(0/7, 0/0, 0/1, 1/2)
Stratum 4		
(0/5, 0/1, 0/3, 1/2)	(0/4, 0/2, 1/3, 0/1)	(0/4, 0/2, 0/2, 1/1)
(0/4, 0/1, 1/4, 0/0)	(0/5, 0/0, 0/2, 1/1)	(0/3, 0/0, 0/2, 1/1)
(0/2, 0/0, 1/1, 0/1)		

The impression is confirmed by the score statistic analysis in Table 2. Most of the evidence for a marker effect derives from Strata 3 and 4 which contain the majority of deaths, though a significant trend test is found even in Stratum 2. The summary chi square values give strong evidence in favor of a marker effect, and most of the variation attributable to CEA state can be explained by the simple trend model with scores 1, 2, 3 and 4. The ratios of observed to expected deaths can be used to obtain a rough idea of the relative hazard. Thus $(8/2.17)/(3/10.73)=13.19$ is a crude estimate of the relative hazard of death for a patient with CEA >10.0 ng/ml compared with that for an otherwise similar patient with CEA $\leqslant 2.5$ ng/ml.

Table 2
Mantel score statistics applied to CEA data

	Omnibus MH test (3 df)	Trend test (1 df)	Observed deaths/expected deaths for covariate level k			
			$k=1$	$k=2$	$k=3$	$k=4$
Stratum 1 alone	1.644	0.405	2/2.45	0/0.09	1/0.36	0/0.11
Stratum 2 alone	—*	5.352	0/0.71	0/0.14	0/0.00	1/0.14
Stratum 3 alone	9.044	7.474	1/4.19	1/0.26	1/0.55	3/0.99
Stratum 4 alone	14.950	12.038	0/3.37	0/0.62	3/2.07	4/0.93
Summary values	24.432	23.000	3/10.73	1/1.12	5/2.98	8/2.17

* Dashes indicate a singular covariance matrix.

Analyses based on the likelihoods (2) and (10) are shown in Table 3. The analysis for grouped data is presented for purposes of illustration only. Such grouped analysis might be misleading because the number of nuisance parameters P_{ij} is so large. Nonetheless, the two analyses agree remarkably well in this example. There is no evidence against the proportional hazards model, and the simple linear trend is seen to fit the data virtually as well as the model with three degrees of freedom. The confidence interval on $\hat{\alpha}_4$ from the ungrouped analysis suggests that the hazard for an individual with CEA >10 ng/ml is between $\exp(1.51)=4.53$ and $\exp(4.53)=92.76$ times that of an individual with CEA $\leqslant$ 2.5 ng/ml. The interpretation of $\hat{\beta}=1.000$ is that the hazard increases by about $\exp(1.000)=2.7$ fold as $Z_1(t)$ values increase from 1 to 2, 2 to 3 or 3 to 4, corresponding respectively to CEA increases from $[0, 2.5]$ to $(2.5, 5]$, from $(2.5, 5]$ to $(5, 10]$, or from $(5, 10]$ to $(10, \infty)$. Note that the likelihood ratio statistics for testing the null hypotheses $\boldsymbol{\alpha}=\mathbf{0}$ or $\beta=0$ are similar to the score statistics in Table 2.

To obtain information on when the CEA marker rises in relation to a subsequent death, a discrete version of the lagged functional $Z_2(t)$ (see §1) was studied. We used the same four discrete levels of CEA as in Tables 1–3, and the results of these analyses are in Table 4. The case of zero lag corresponds to Tables 1–3. It is seen that the unadjusted analyses that ignore stratifying variables give even more striking evidence for a marker effect than the adjusted analyses, a fact to be discussed in §4. As the lag time increases it is seen that the score statistics decrease and values of $\hat{\alpha}_2$ and $\hat{\alpha}_3$ become smaller and variable in sign. However, $\hat{\alpha}_4$ is roughly 2, even for a lag of 300 days, and the corresponding score tests remain statistically significant. This analysis suggests that a CEA elevation above 10 ng/ml, measured three to ten months previously, carries an increased risk of about $\exp(2)=7$-fold, compared to an increased risk of $\exp(3.02)=20$-fold for a current CEA exceeding 10 ng/ml. Thus, if a patient has had a CEA exceeding 10 ng/ml, even if the measurement was obtained as much as ten months earlier, he is at increased risk of death. Analysis of the functional $Z_3(t)$ defined in §1 showed that a patient who has ever had a CEA exceeding 10 ng/ml had three times the risk of a patient who had never had such an elevation (the score test statistic was 9.53 with 1 df).

Time to recurrence (instead of death) was also examined. Analysis as in Tables 2 and 3 yielded $\hat{\alpha}_2=0.26$, $\hat{\alpha}_3=1.09$, $\hat{\alpha}_4=1.92$, a 3 df score of 19.90 and a 1 df trend score of 18.52. However, examination of lagged CEA levels as in Table 4 revealed no impressive marker effects for lags of 100 days or more (values of $\hat{\alpha}_4$ were 1.61, 0.71 and 1.52 for lags of 100, 200 and 300 days, but tests of $\boldsymbol{\alpha}=\mathbf{0}$ were not statistically significant). Thus current CEA elevation is associated with current increased risk of recurrence, but the data are less

Table 3
Likelihood analysis of CEA data

Following methods for continuous time to response			Following methods for grouped data		
Parameter estimates and 95% confidence intervals					
$\hat{\alpha}_2$	1.164	(−1.14, 3.47)	1.195	(−1.13, 3.52)	
$\hat{\alpha}_3$	2.103	(0.50, 3.71)	2.301	(0.66, 3.94)	
$\hat{\alpha}_4$	3.019	(1.51, 4.53)	3.391	(1.83, 4.95)	
$\hat{\beta}$	1.000	(0.51, 1.49)	1.140	(0.63, 1.65)	
Estimated covariance of $(\hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)'$					
1.380	0.341	0.337	1.403	0.347	0.345
	0.673	0.417		0.700	0.439
		0.592			0.634
Estimated variance of $\hat{\beta}$					
0.0623			0.0690		

Tests of the model

Source	SS	df	Source	SS	df
None	—	—	Proportional hazards model (1)	35.84	48
Deviations from linear trend model (5)	0.04	2	Deviations from linear trends model (5)	0.01	2
None	—		Linear trend model	35.85	50

Likelihood ratio tests of null hypotheses $\boldsymbol{\alpha} = 0$ and $\beta = 0$

20.68 (3 df)	24.97 (3 df)
20.64 (1 df)	24.96 (1 df)

Table 4
Analyses of lagged CEA levels

Adjusted for stratum variables							Unadjusted					
Lag (days)	Valid deaths	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	3 df score test	1 df trend test	Valid deaths	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	3 df score test	1 df trend test
0	20	1.16	2.10	3.02	24.43	23.00	20	2.07	2.59	3.70	64.59	58.78
100	24	1.20	−0.73	2.02	17.43	10.49	24	0.85	−0.60	2.32	33.65	15.37
200	20	−0.19	0.72	2.28	11.35	7.49	20	−0.31	1.06	1.71	11.57	9.29
300	16	0.13	0.61	1.99	12.56	6.15	16	−0.75	0.45	2.09	8.06	5.54

clear regarding lagged CEA levels. Such a finding is in line with the notion that CEA rises about the same time that other signs of recurrence become evident, and this happens months or years before the patient succumbs.

Moertel *et al.* (1978) found CEA levels above 5 ng/ml in only nine of 27 patients with recurrent disease and in none of nine with local residual disease after resection. They conclude that the 'CEA assay is not a sensitive test for early detection of recurrence of colorectal carcinoma', and this is literally so, in view of the small proportion with elevated CEA. Our methods do not address test sensitivity, which is an important aspect of clinical utility. Rather, we seek to determine only whether those patients who do have marker elevations are at risk. Our finding that current CEA elevation, however rarely it occurs, carries increased risk of recurrence and death is an important first step in establishing clinical utility. But the fact that we were unable to demonstrate that previous CEA elevations carry increased current risk of recurrence (though they do carry increased current risk of death), and the low test sensitivities observed by Moertel *et al.* (1978) support the pessimistic assessment of those authors.

4. Discussion

Serial marker data are complicated by censorship of the response variable, such as recurrence time, as well as by incomplete knowledge of the patient's marker state. Both these difficulties yield to partial likelihood methods based on the proportional hazards model with time-dependent covariate because individuals are members of the 'risk set' (see Cox, 1972) or not according as they are alive, under follow-up, and of known marker state or not at the death time which corresponds to that risk set. Recent independent work by Prentice *et al.* (1978) outlines a similar approach to the problem relating the extent of graft versus host disease to risk of leukemia recurrence. It is likely that such methods will prove useful for a wide variety of markers and diseases.

We have assumed that the timing and measurement procedures for the marker are not dependent on patient status. Clearly, if the investigator stops taking measurements as the patient sickens, or if the technician is notified when a recurrence has occurred, such investigations may fail. Ideally, such measurements are made at frequent scheduled intervals by a technician who is not informed of the patient's status.

Lack of frequent marker measurements can seriously degrade the power of marker studies. In the CEA example, only 20 of the 34 deaths were valid, and only 17 deaths were informative. These figures are based on the functional $\mathbf{Z}(t)$ in Table 1 and the convention that $\gamma(t)$ was censored at t unless the most recent measurement was no more than 100 days before. The Fisher information is essentially proportional to the number of informative deaths. To see this consider the continuous time case with scalar α. Then, an estimate of the Fisher information is $\sum pq$ where $p \equiv N_1 \exp(\alpha)/\{N_1 \exp(\alpha) + N_2\}$, N_1 and N_2 are the numbers at risk in states $Z(t) = 1$ and $Z(t) = 0$ within a given stratum, and the summation extends over all informative death times. If N_1 and N_2 are always large compared to the number of informative death times, the information is seen to be essentially proportional to the number of terms in the summation.

Information may also be lost by overstratification. In Table 4, ignoring the stratum variables leads to an increase in the score statistic test of $\boldsymbol{\alpha} = 0$ from 24.43 to 64.59. If, in the previous information calculation, the numbers at risk N_1 and N_2 within each stratum are large compared to the number of informative deaths, then the information is essentially independent of the stratification plan and depends mainly on the number of such deaths. If, however, stratification is so extensive that the numbers at risk N_1 and N_2

within a stratum are comparable to the numbers of informative deaths in that stratum, then, for $\exp(\alpha) > 1$, the numbers at risk in state $Z(t) = 1$ will tend to be depleted so that the information pq from deaths after depletion will diminish, and in the cases $N_1(t) = 0$, which correspond to a valid but uninformative death, will equal zero. If previous work demonstrates the importance of certain stratifying factors, one may be obligated to perform a stratified analysis, even at the risk of overstratification. However, if attention to such factors really improves the fit of the model, and if overstratification can be avoided, power should be increased by such adjustment.

At the risk of an added assumption, one can adjust for stratum effects by using the model $h(t) = h_0(t) \exp\{\boldsymbol{\alpha}'\mathbf{Z}(t) + \beta_i\}$ where β_i is a stratum effect. This model assumes that all strata have proportional hazards, which is not required by (1). Under this model, overstratification is avoided in the previous example, and the 3 df likelihood ratio test of $\boldsymbol{\alpha} = 0$ increases from 20.68 under (1) to 41.72. More generally one could adjust for other covariates through a model such as $h(t) = h_0(t) \exp\{\boldsymbol{\alpha}'\mathbf{Z}(t) + \boldsymbol{\beta}'\mathbf{U}(t)\}$, where some of the components of $\mathbf{U}(t)$ might be time-dependent and others might be initial covariates.

Acknowledgements

I wish to thank Drs T. M. Chu, E. D. Holyoke and P. Lavin for providing the data in Table 1, and gratefully acknowledge the help of Mrs Marie Topor for computing the risk sets and covariate levels in that example. I also thank Drs David Byar and Sylvan Green and *Biometrics* reviewers for helpful comments, and I thank Miss Julie Paolella for secretarial help.

Résumé

Les modèles de hasard proportionnels de Cox (1972, *Journal of the Royal Statistical Society, Séries B* **34,** 187–220), avec une covariable dépendant du temps, sont utilisés pour l'analyse de données de marqueur de cancer. Un avantage particulier de cette méthode est la facilité avec laquelle les données manquantes sont traitées. L'analyse d'une suite de données réelles montre que de hauts niveaux de marqueurs de cancer carcinoembryoniques antigen (CEA) sont associés à un risque de mort croissant chez des patients avec un cancer colorectal réséqué. Plusieurs données décrivant l'évolution du CEA sont analysées: le niveau de CEA au temps de mort t; le niveau de CEA 200 jours avant le temps t; le niveau de CEA dépasse-t-il 5 ng/ml avant t? Les méthodes décrites permettent de tester les hypothèses 'pas d'effet marqueur', de donner des estimateurs et des intervalles de confiance pour les paramètres du modèle, à la fois pour des résponses continues et pour des résponses groupées dans le temps. Pour les données groupées, un test du rapport de vraisemblance pour l'hypothèse de hasards proportionnels est suggéré.

References

Booth, S. N., Jamieson, G. C., King, J. P. G., Leonard, J., Oates, G. D. and Dykes, P. W. (1974). Carcinoembryonic antigen in the management of colorectal carinoma. *British Medical Journal* **4,** 183–187.

Breslow, N. E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review* **43,** 45–57.

Chu, T. M., Holyoke, E. D., Cedermark, B., Evans, J. and Fischer, D. (1976). A long-term follow-up of CEA in resective colorectal cancer. In *Onco-Developmental Gene Expression*, W. H. Fishman and S. Sell (eds), 427–431. New York: Academic Press.

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

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72,** 27–36.

Gail, M. (1975). A review and critique of some models used in competing risk analysis. *Biometrics* **31,** 209–222.

Mach, J. P., Jaeger, Ph., Bertholet, M. M., Ruegsegger, C. H., Loosli, R. M. and Pettavel, J. (1974). Detection of recurrence of large bowel carcinoma by radioimmunoassay of circulating carcinoembryonic antigen (C.E.A.). *Lancet* **2,** 535–540.

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. and 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. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22,** 719–748.

Moertel, C. G., Schutt, A. J. and Go, V. L. W. (1978). Carcinoembryonic antigen test for recurrent colorectal carcinoma: Inadequacy for early detection. *Journal of the American Medical Association* **239,** 1065–1066.

Pastorfide, G. B., Goldstein, D. P. and Kosasa, T. S. (1974). The use of a radioimmunoassay specific for human chorionic gonadotropin in patients with molar pregnancy and gestational trophoblastic disease. *American Journal of Obstetrics and Gynecology* **120,** 1025–1028.

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34,** 57–67.

Prentice, R. L., Kalbfleisch, J. D., Peterson, Jr, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34,** 541–554.

Tarone, R. E. (1975). Tests for trend in life table analysis. *Biometrika* **62,** 679–682.

Waldmann, T. A. and McIntire, R. K. (1974). Use of radioimmunoassay for alpha fetoprotein in the diagnosis of malignancy. *Cancer* **34,** 1510–1515.

Received March 1979; *revised February and June* 1980

Appendix

Formulas for Grouped Data

To maximize (10), one can solve the following two sets of equations iteratively:

$$\sum_{k=1}^{K} \{S_{ijk} - D_{ijk}\hat{\psi}_{ijk}(1-\hat{\psi}_{ijk})^{-1}\} \exp(\hat{\alpha}_k) = 0 \tag{A1}$$

for all i and j, and

$$\sum_{i,j} \{S_{ijk} - D_{ijk}\hat{\psi}_{ijk}(1-\hat{\psi}_{ijk})^{-1}\} \ln \hat{P}_{ij} = 0 \tag{A2}$$

for $k = 2, 3, \ldots, K$. In these equations, $\hat{\psi}_{ijk} = \hat{P}_{ij} ** \exp(\hat{\alpha}_k)$.

One can estimate the covariance of $\boldsymbol{\alpha}' = (\hat{\alpha}_2, \hat{\alpha}_3, \ldots, \hat{\alpha}_k)$ from the information matrix $\mathbf{I}$ associated with (10). The matrix $\mathbf{I}$ can be partitioned into the $(K-1)\times(K-1)$ diagonal matrix $\mathbf{I}_{11}$ associated with $\hat{\boldsymbol{\alpha}}$, the $L \times L$ diagonal matrix I_{22} associated with $(\hat{P}_{11}, \ldots, \hat{P}_{1J_1}, \ldots, \hat{P}_{IJ_I})$, and the remaining matrices $\mathbf{I}_{12} = \mathbf{I}_{21}'$. The kth diagonal element of $\mathbf{I}_{11}$ is estimated as

$$\sum_{i,j} \ln \hat{P}_{ij} \exp(\hat{\alpha}_k)[D_{ijk}\hat{\psi}_{ijk}(1-\hat{\psi}_{ijk})^{-1}\{1+\exp(\hat{\alpha}_k)\ln \hat{P}_{ij}(1-\hat{\psi}_{ijk})^{-1}\} - S_{ijk}]. \tag{A3}$$

The lth diagonal element of $\mathbf{I}_{22}$, corresponding to stratum i and interval j, has estimate

$$\sum_{k} \hat{P}_{ij}^{-2} \exp(\hat{\alpha}_k)[S_{ijk} + D_{ijk}\hat{\psi}_{ijk}\{\exp(\hat{\alpha}_k) - 1 + \hat{\psi}_{ijk}\}\{1-\hat{\psi}_{ijk}\}^{-2}]. \tag{A4}$$

The (k, l) element of $\mathbf{I}_{12}$ is estimated by

$$\hat{P}_{ij}^{-1} \exp(\hat{\alpha}_k)[D_{ijk}\hat{\psi}_{ijk}\{1-\hat{\psi}_{ijk} + \exp(\hat{\alpha}_k)\ln \hat{P}_{ij}\}(1-\hat{\psi}_{ijk})^{-2} - S_{ijk}]. \tag{A5}$$

The desired covariance estimate is

$$\hat{\boldsymbol{\Sigma}}_{11} = (\hat{\mathbf{I}}_{11} - \hat{\mathbf{I}}_{12}\hat{\mathbf{I}}_{22}\hat{\mathbf{I}}_{21})^{-1}. \tag{A6}$$

In a similar fashion, one can maximize the grouped likelihood based on the trend model (5) by

solving the following equations iteratively with $\hat{\Gamma}_{ijk} = \hat{P}_{ij} * * \exp(\hat{\beta} z_k)$:

$$\sum_k \{S_{ijk} - D_{ijk}\hat{\Gamma}_{ijk}(1-\hat{\Gamma}_{ijk})^{-1}\} \exp(\hat{\beta} z_k) = 0 \tag{A7}$$

for all i and j, and

$$\sum_{i,j,k} \{S_{ijk} - D_{ijk}\hat{\Gamma}_{ijk}(1-\hat{\Gamma}_{ijk})^{-1}\} z_k \ln \hat{P}_{ij} \exp(\hat{\beta} z_k) = 0. \tag{A8}$$

An estimate of the scalar I_{11} is then

$$\sum_{i,j,k} z_k^2 \ln \hat{P}_{ij} \exp(\hat{\beta} z_k)[D_{ijk}\hat{\Gamma}_{ijk}(1-\hat{\Gamma}_{ijk})^{-1}\{1+\ln \hat{P}_{ij} \exp(\hat{\beta} z_k)(1-\hat{\Gamma}_{ijk})^{-1}\} - S_{ijk}], \tag{A9}$$

and an element of the $1 \times L$ matrix $\hat{\mathbf{I}}_{12}$ is

$$\sum_k z_k \hat{P}_{ij}^{-1} \exp(\hat{\beta} z_k)[D_{ijk}\hat{\Gamma}_{ijk}\{1-\hat{\Gamma}_{ijk} + \ln \hat{P}_{ij} \exp(\hat{\beta} z_k)\}(1-\hat{\Gamma}_{ijk})^{-2} - S_{ijk}]. \tag{A10}$$

The nonzero elements of the diagonal $L \times L$ matrix $\hat{\mathbf{I}}_{22}$ are obtained from (A4) with $\hat{\beta} z_k$ replacing $\hat{\alpha}_k$. One may test the model (5) using (14) with $\hat{\beta} z_k$ replacing $\hat{\alpha}_k$ and with $L(K-1)-1$ degrees of freedom.