Biometrika (1994), **81**, 3, *pp.* 501–14
Printed in Great Britain

A partly parametric additive risk model

BY IAN W. MCKEAGUE
Department of Statistics, Florida State University, Tallahassee, Florida 32306-3033, U.S.A.

AND PETER D. SASIENI
Department of Mathematics, Statistics and Epidemiology, Imperial Cancer Research Fund, P.O. Box No. 123, Lincoln's Inn Fields, London WC2A 3PX, U.K.

SUMMARY

Aalen's additive risk model allows the influence of each covariate to vary separately over time. Although allowing greater flexibility of temporal structure than a Cox model, Aalen's model is more limited in the number of covariates it can handle. We introduce a partly parametric version of Aalen's model in which the influence of only a few covariates varies nonparametrically over time, and that of the remaining covariates is constant. Efficient procedures for fitting this new model are developed and studied. The approach is applied to data from the Medical Research Council's myelomatosis trials.

Some key words: Aalen's linear hazards model; Counting process; Efficient estimation; Right-censored data; Semiparametric; Survival analysis.

1. INTRODUCTION

The additive hazards model of Aalen (1980) has received relatively little attention. Judged in terms of achieving a reasonable fit to data, this model should perform well since it is the first step of a Taylor series expansion of a general hazard function about the zero of the covariate vector. However, in estimating the unknown functions in such a general model there is a variance-bias trade-off that may be critical in small and medium samples. Also, after fitting the model one does not have parameters or formulae that are easily reported. We propose a model that takes the additive structure of Aalen's model and imposes parametric constraints to obtain a semiparametric submodel, which may be more appropriate in some applications.

The model will be illustrated with data from clinical trials on myelomatosis. Covariates include treatment, sex and four age strata, which will be treated parametrically, together with serum levels of haemoglobin and β_2-microglobulin, whose effects will be investigated nonparametrically. The additive form can be interpreted loosely in terms of unobserved competing risks since the hazard function for the minimum of independent random variables is the sum of the hazard functions for the individual variables. Microglobulin levels are related to kidney function and tumour mass, whereas haemoglobin is unaffected by kidney function. Hence one might anticipate that the hazard function associated with each covariate represents a different cause of death.

We model the hazard at time t by

$$\lambda(t|x, z) = \alpha(t)'x + \beta'z, \tag{1·1}$$

where x and z are q and p dimensional covariates respectively, and $\alpha(.)$ and β are unknown.

The influence of the covariates in x can vary with time, but that of z is restricted to be constant. By defining a new time dependent covariate, $z^*(t) = z \exp(-t)$ say, nonconstant time effects could be included. The first component of x may be set to 1 to allow for a general baseline hazard. We are interested in estimating β, the vector of 'cumulative hazards'

$$A(.) = \int_0^{\cdot} \alpha(s)\, ds,$$

and the conditional survival function for given values of the covariates.

Recent work on estimation in Aalen's (1980) model $\lambda(t|x) = \alpha(t)'x$ has been carried out by Huffer & McKeague (1991) and McKeague (1988a, b). Greenwood & Wefelmeyer (1990, 1991) and Sasieni (1992) have shown that the Huffer–McKeague estimator is asymptotically efficient and that it is an approximate maximum likelihood type estimator. The 'full Aalen' model has had only limited use in data analysis, primarily in data sets with just a few covariates. Examples are given by Aalen (1989, 1993), Mau (1986, 1988) and Henderson & Milner (1991). Lin & Ying (1994), among others, have considered an additive analogue of Cox's (1972) proportional hazards model:

$$\lambda(t|z) = \alpha_0(t) + \beta' z, \tag{1·2}$$

which is considerably less versatile than (1·1).

In § 2 we derive estimators for β and A and discuss their practical implementation. Section 3 considers model selection and gives an influence function diagnostic residual. The application to finding prognostic factors for survival among myelomatosis patients is discussed in § 4. Some general discussion, comparing the new approach with the standard Cox model approach, is provided in § 5. FORTRAN computer code can be obtained from the authors.

2. SEMIPARAMETRIC ESTIMATORS

2·1. *General*

If β were known one could use Aalen's least squares estimator for $A(.)$. Similarly, if $\alpha(.)$ were known one could estimate β by maximum likelihood. Intuitively, iterating between estimation of β and α should work. Here we use the score equation for β to derive a set of pseudo-normal equations. There are similarities with the approach used by Sasieni (1992) to motivate estimators for the full Aalen model. The first step is to derive ordinary least squares estimators. If ease of calculation is important, these are the most appropriate estimators. They are unbiased, consistent and asymptotically normal. Efficient estimators can be obtained via weighted least squares, which requires smoothing of a preliminary estimate of A to obtain the weights.

2·2. *Ordinary least squares estimators*

Denote by $(x_i, z_i, T_i, \delta_i)$ the observed covariates x_i and z_i, possibly censored failure time T_i, and censoring indicator δ_i, for the ith of n individuals, from independent and identically distributed copies of a generic (x, z, T, δ); $\delta_i = 1$ if T_i is uncensored. Recall that a density f is related to its hazard λ by

$$f(t) = \lambda(t) \exp\left\{ - \int_0^t \lambda(u)\, du \right\}.$$

Thus, in the usual survival set-up with noninformative and conditionally independent censoring, the log-likelihood for λ is

$$\sum_{i=1}^{n} \left\{ \delta_i \log \lambda_i(T_i) - \int 1_{[T_i \geqslant t]} \lambda_i(t)\, dt \right\}, \tag{2·1}$$

where the range of integration extends over the follow-up period and $\lambda_i(t) = \lambda(t | x_i, z_i)$ (Kalbfleisch & Prentice, 1980, eqn (5·2)). Assume that $\lambda(. | x, z)$ is bounded away from zero.

Differentiate (2·1) with respect to β to obtain the parametric score function:

$$\begin{aligned} \dot{l}_\beta &= \sum_{i=1}^{n} \left\{ \delta_i \frac{z_i}{\lambda_i(T_i)} - \int 1_{[T_i \geqslant t]} \frac{z_i}{\lambda_i(t)} \lambda_i(t)\, dt \right\} \\ &= \sum_{i=1}^{n} \left\{ \delta_i \frac{z_i}{\lambda_i(T_i)} - \int 1_{[T_i \geqslant t]} \frac{z_i z_i'}{\lambda_i(t)} \beta\, dt - \int 1_{[T_i \geqslant t]} \frac{z_i x_i'}{\lambda_i(t)} \alpha(t)\, dt \right\}. \end{aligned}$$

Setting $\dot{l}_\beta = 0$ yields

$$\beta = \left(\int Z'WZ\, dt \right)^{-1} \left(\int Z'W\, dN - \int Z'WX\, dA \right), \tag{2·2}$$

where $Z = Z(t) = (z_1 1_{[T_1 \geqslant t]}, \ldots, z_n 1_{[T_n \geqslant t]})'$, X is defined like Z, the $n \times n$ matrix W is given by $W(t) = \operatorname{diag}\{1/\lambda_i(t)\}$, $N(t) = (N_1(t), \ldots, N_n(t))'$ and $N_i(t) = 1_{[T_i \leqslant t, \delta_i = 1]}$ is the process that counts an uncensored failure of individual i.

Next consider a submodel, $\alpha(t) = \alpha(t; \eta) = \alpha_0(t) + \eta b(t)$, in which η is a one-dimensional parameter and b is a given q-vector of functions. Differentiating (2·1) with respect to η gives

$$\dot{l}_\eta = \dot{l}_\alpha b = \int b'X'W\, dN - \int b'X'WZ\beta\, dt - \int b'X'WX\, dA.$$

We are not only interested in the particular submodel, but in all such models. More rigorously, consider a family of regular parametric submodels whose closure is the semiparametric model. Since the parametric models are special cases of (1·1), an estimator for the general model should work on all submodels. Setting $\dot{l}_\alpha b = 0$ for all vector valued functions b implies that

$$A(t) = \int_0^t (X'WX)^{-1}(X'W\, dN - X'WZ\beta\, ds). \tag{2·3}$$

Substituting the right-hand side of (2·3) into (2·2) and solving for β gives

$$\hat{\beta} = \left(\int Z'HZ\, dt \right)^{-1} \int Z'H\, dN, \tag{2·4}$$

where $H = W - WX(X'WX)^{-1}X'W$. But $\hat{\beta}$ is not an estimator since it depends on the unknown λ. However, replacing W in H by the identity matrix I yields an estimator $\tilde{\beta}$ of β which is analogous to Aalen's (1980) least squares estimator. It is exactly unbiased, $\sqrt{n}$-consistent and asymptotically normal. Replacing β in (2·3) by $\tilde{\beta}$ and using I in place of W gives an unbiased, $\sqrt{n}$-consistent and asymptotically Gaussian estimator of A. To calculate the estimators, assume that X and Z are constant between failure times, so that integration can be replaced by summation, and replace the terms 'ds' by $T_{(i)} - T_{(i-1)}$, where

the $T_{(i)}$ are the ordered failure times. Integration with respect to the counting process is by definition simply summation.

2·3. *Weighted least squares estimators*

We now construct efficient estimators for β and A. In parametric models with a parameter of interest θ and a nuisance parameter ϕ, the efficient score for θ is

$$l_\theta^* = \dot{l}_\theta - E[\dot{l}_\theta \dot{l}_\phi'] \{E[\dot{l}_\phi \dot{l}_\phi']\}^{-1} \dot{l}_\phi,$$

which is a function of θ and ϕ in general. If $\hat{\theta}$ and $\tilde{\phi}$ are consistent estimators of θ and ϕ satisfying $l_\theta^*(\hat{\theta}, \tilde{\phi}) = 0$, then $\hat{\theta}$ is asymptotically efficient. The same ideas carry over to semiparametric models. In our model, $\hat{\beta}$ solves $\hat{l}_\beta^* = 0$, where $\hat{l}_\beta^*$ is an approximate efficient score for β, as shown in Appendix 2. An estimator based on (2·4), but with a consistent estimate of λ replacing the unknown function, turns out to be efficient for the semiparametric model (Bickel et al., 1993, § 3.4).

We propose two methods of constructing efficient estimators by replacing W by $\hat{W} = \text{diag}\{1/\hat{\lambda}_i(.)\}$, where $\hat{\lambda}_i$ is some estimate of λ_i. The second is more appropriate when the dimension of β is large.

Method 1

(i) Fit the Aalen model, $\lambda(t|x, z) = \alpha(t)'x + \beta(t)'z$, and obtain $\hat{W}$ from a predictable kernel smoother, following Huffer & McKeague (1991).
(ii) Find an estimate $\hat{\beta}$ of β by (2·4), using $\hat{W}$ in place of W.
(iii) Estimate A from (2·3) using $\hat{W}$ and $\hat{\beta}$ in place of W and β.

Method 2

(i) Obtain initial estimates of β and A by ordinary least squares, as in § 2·2.
(ii) Use a predictable kernel smoother to estimate α.
(iii) Obtain $\hat{W}$ using the estimates of β and α from (i) and (ii).
(iv) Obtain final estimates $\hat{A}$ and $\hat{\beta}$ using (2·3) and (2·4) with $\hat{W}$ in place of W.

We have used Method 2 with a $\hat{\lambda}_i$ explicitly defined in § 2·6. Notice that Z and X are functions of t and the same estimating equations (2·3) and (2·4) could be used with predictable time-dependent covariates.

The gain in efficiency using $\hat{W}$ compared to I will depend on the heterogeneity of the hazards in the sample. If all individuals are at equal risk, so that none of the covariates are related to survival, there is no efficiency gain. In general, however, there will be a small gain. Huffer & McKeague (1991) investigated by simulation the asymptotic relative efficiency of the ordinary least squares estimator in the Aalen model and found it to be between 72% and 98% depending on the distribution of the covariates and the magnitude of the risk associated with them. The situation is more complicated here because $\hat{\beta}$ depends on the weights for all individuals at risk at each failure time. For a given data set, the efficiency gain can be examined by comparing the estimated asymptotic covariance matrices, or by a bootstrap simulation.

2·4. *Estimating the asymptotic covariance matrix*

The asymptotic distribution of $\hat{\beta}$ and $\hat{A}$ is the same for Methods 1 and 2. Counting process techniques (Appendix 1) can be used to show that $n^{\frac{1}{2}}(\hat{\beta} - \beta)$ converges in distribution to a p-variate normal with mean zero and with covariance matrix which can be

consistently estimated by $\hat{\Sigma}^{-1}$, where $\hat{\Sigma} = n^{-1}\int Z'\hat{H}Z\,dt$. Here $\hat{H}$ is obtained by replacing W by a consistent estimate $\hat{W}$ in H. Further, $n^{\frac{1}{2}}(\hat{A} - A)$ converges in distribution to a q-variate Gaussian process with mean zero and covariance function which, as a function of s and t, can be consistently estimated by

$$n\sum_{u \leqslant s \wedge t} \Delta_u \Delta_u' + \hat{\psi}(s)\hat{\Sigma}^{-1}\hat{\psi}(t)', \tag{2·5}$$

where Δ_u is the jump in $\hat{A}$ at time u and $\hat{\psi}(t) = \int_0^t (X'\hat{W}X)^{-1}X'\hat{W}Z\,ds$. The first term in (2·5) is a consistent estimate of the covariance function for the model having only nonparametric terms; the second represents the contribution from the parametric part.

2·5. *Grouped data version*

One may wish to fit a grouped data version of (1·1) in the exploratory stage of model building. For most purposes grouping the time axis into ten intervals would be adequate and greatly reduce computation. The grouped model may be written

$$\lambda(t\,|\,x, z) = \sum_{j=1}^{K} \alpha_{(j)}' x 1_{\mathscr{I}_j}(t) + \beta' z,$$

where the intervals are $\mathscr{I}_j = [\tau_{j-1}, \tau_j)$, for $j = 1, \ldots, K$, and $\tau_0 = 0$. One approach treats this as a parametric linear model with $Kq + p$ parameters, but it makes sense to take into account the orthogonality of the dummy covariate blocks $x1_{\mathscr{I}_j}(t)$ for $j = 1, \ldots, K$. Let $k(u)$ denote the index such that $u \in \mathscr{I}_{k(u)}$. Using the same argument as was used to derive (2·2) and (2·3), one has

$$A(u) = \sum_{j=1}^{k(u)-1} \alpha_{(j)}(\tau_j - \tau_{j-1}) + \alpha_{k(u)}(u - \tau_{\{k(u)-1\}}),$$

$$\alpha_{(j)} = \left(\int_{\mathscr{I}_j} X'WX\,dt\right)^{-1} \int_{\mathscr{I}_j} (X'W\,dN - X'WZ\beta\,du).$$

Thus, instead of having to solve a system of q linear equations at each failure time, one only has to solve such a system for each of the K intervals.

2·6. *The choice of weights*

For the asymptotic theory $\hat{\lambda}_i$ need only be consistent, but in practice the choice of $\hat{\lambda}_i$ requires some care. It is a good idea to compare the weighted with the unweighted estimates of β and A, since both are consistent at the model, and disagreement may indicate that the model or the weights are inappropriate.

We calculated $\hat{\lambda}_i$ as follows. Let $T_{(i)}$ denote the ith ordered failure time, and set $T_{(0)} = 0$. Given the initial estimate $\tilde{A}$ of A, estimate $\alpha(t)$ for $t > T_{(d)}$ by

$$\tilde{\alpha}(t) = \frac{\tilde{A}(T_{(i)}) - \tilde{A}(T_{(i-d)})}{T_{(i)} - T_{(i-d)}}$$

when $T_{(i)} < t \leqslant T_{(i+1)}$. Taking d between $n^{\frac{1}{2}}$ and $4n^{\frac{1}{2}}$ works well for n between 100 and 1000. Notice that $\tilde{\lambda}_i(t) = \tilde{\alpha}(t)'x_i + \tilde{\beta}'z_i$ estimates $\lambda_i(t)$, but it may be nonpositive and it is undefined for $t \leqslant T_{(d)}$. Instead, we use

$$\hat{\lambda}_i(t) = \begin{cases} \max\{\varepsilon\bar{\lambda}(t), \tilde{\lambda}_i(t)\} & \text{for } t > T_{(d)}, \\ \bar{\lambda}(t) & \text{for } t \leqslant T_{(d)}, \end{cases}$$

where $\bar{\lambda}(t)$ is the average of $\tilde{\lambda}_i(t \vee T_{(d+1)})$ over all individuals i at risk at time $t-$. We recommend taking ε between 0·15 and 0·35. In our examples, $d = 50$ and $\varepsilon = 0{\cdot}25$. Strictly speaking, $\hat{\lambda}_i$ departs from Methods 1 and 2 for $t \leqslant T_{(d)}$, but this will have negligible effect provided that d is small compared to the total number of observed failures. Many variations on this recipe for $\hat{\lambda}_i$ are possible. For instance, the 'bandwidth' for estimating $\hat{\alpha}$ could be taken to be a fixed length of time or a fixed number of uncensored failures.

3. MODEL REFINEMENT AND DIAGNOSTICS

When should a covariate be treated nonparametrically? There may be strong scientific reasons to do so with some covariates, but initially it is advisable to treat at most a few that way. Yet a factor may not be significant when modelled parametrically even though it has a strong effect on survival, e.g. a drug that is strongly toxic, but which helps those patients who survive the initial period of toxicity. It is generally sensible to include a nonparametric baseline. By centring the covariates, the baseline can have a meaningful interpretation as the hazard for an 'average' individual. To examine whether the influence of the parametric covariates varies with time, treat each nonparametrically and look at the plot of $\hat{A}_j(t)$ along with the corresponding straight line estimate $t\hat{\beta}_j$. These plots together with pointwise confidence intervals will give some indication of the validity of the parametric assumptions and how they are violated when they fail. This approach is illustrated in § 4.

Other approaches are possible. One might fit a separate Aalen model for each covariate, to get an initial idea of the variation of the hazard with time, before attempting multivariate modelling. Alternatively, after selecting a partly parametric model, one might check to see if any of the variables not included make a significant nonparametric contribution.

Influence residuals approximate the change in estimates when an observation is removed. For fixed W, $\hat{\beta}$ as defined in (2·3) is an explicit functional of the empirical distribution. Differentiating this functional and evaluating the derivative at the empirical distribution gives the empirical influence curve (Cook & Weisberg, 1982, pp. 104–8). Straightforward differentiation and a little algebra yields

$$\partial\hat{\beta}_i = \left(\int Z'HZ\, dt \right)^{-1} \int \{z_i - Z'WX(X'WX)^{-1}x_i\} W_{ii}\, d\hat{M}_i$$

as the influence of the ith individual on $\hat{\beta}$. Here

$$\hat{M}_i(t) = N_i(t) - \int_0^t 1_{[T_i \geqslant s]}(x_i' d\hat{A} + \hat{\beta}' z_i\, ds)$$

is the ith martingale residual. The effect of estimating W on the influence curve for β is asymptotically negligible whenever the assumed model holds. That is, if the influence curve is evaluated at a probability measure for a partly additive Aalen model (1·1) then the above expression for the $\partial\hat{\beta}_i$'s will be correct to first order. Henderson & Oman (1993) study influence curves for the full Aalen model.

The basic diagnostic building block is the counting process martingale residual $\hat{M}_i(.)$; see Barlow & Prentice (1988), Therneau, Grambsch & Fleming (1990) and compare with Aalen (1993). A plot of $\sum_i \hat{M}_i(t)$ against t can be used to check for lack of fit due to components not being allowed to vary freely in time, as in (1·2). To investigate the role of individual covariates, partition the time axis into about ten intervals $[\tau_{j-1}, \tau_j)$ and plot

the increments $\hat{M}_i(\tau_j) - \hat{M}_i(\tau_{j-1})$ against covariates for individuals at risk at τ_{j-1}. Such 'partial residual plots' may detect the need to transform a covariate. To check whether the additive risk associated with a given covariate varies in time, compare the plots for that covariate, after rescaling each by $\tau_j - \tau_{j-1}$.

4. Example

Here the partly parametric additive risk model is applied to data from the Medical Research Council's (1984) fourth and fifth myelomatosis trials. We begin by analyzing data from the fourth trial using just two covariates, one of which is treated nonparametrically. The full Aalen model could be applied with just two covariates, but our model fits the data adequately. For the fifth trial we used seven covariates, only one of which entered nonparametrically. The full Aalen model would be less suitable in that case.

From the fourth trial we analyzed survival data on 495 myelomatosis patients for whom presentation measurements included serum β_2-microglobulin (s-β_2m) in mg/l and serum haemoglobin (Hb) in g/l. Percentiles of these measurements are given in Table 1. In fitting the regression models, s-β_2m was transformed by $\log_{10}(.)$, to compensate for its skewness, and then centred by 0·6. Haemoglobin was centred by 100.

Table 1. *Percentiles of serum β_2-microglobin* (s-β_2m) *and haemoglobin* (Hb)

Covariate	min	10	25	50	75	90	max
s-β_2m	0·3	2·3	3·3	5·7	9	22	76·7
Hb	25	71	90	106	122	136	167

Several studies have indicated that s-β_2m is of primary importance in predicting survival in myelomatosis patients. However, Cuzick et al. (1990) suggest that its value is confined to the first two years of follow-up. This claim was based on an analysis using separate proportional hazards models for different follow-up intervals. Such an approach has limited ability to model covariate effects that vary in their influence over time. It is more appropriate to use a partly parametric additive risk model when searching for such variations.

We initially treated both covariates parametrically and the baseline nonparametrically, as in (1·2). The Wald statistics were 2·25 for s-β_2m and $-3{\cdot}24$ for haemoglobin, strongly suggesting that both covariates are influential. Next we treated haemoglobin parametrically, and the baseline hazard and s-β_2m nonparametrically. This was our final model. Figure 1 shows plots of the cumulative risks for the two nonparametric terms; Fig. 1(b) also contains the straight line estimate of the cumulative risk for s-β_2m based on model (1·2). Note that in the first three years the straight line falls outside the 95% pointwise confidence limits, strongly suggesting that the influence of s-β_2m varies with time. The plateau in cumulative risk after two years in Fig. 1(b) is consistent with the claim of Cuzick et al. (1990) that s-β_2m is of primary importance in predicting survival only within the first two years of follow-up. Haemoglobin treated parametrically has significant influence: Wald $= -3{\cdot}13$. Figure 2 shows that its influence does not vary appreciably since the straight line estimate of cumulative risk is almost completely within the 95% pointwise confidence limits around the Aalen model estimate.

The confidence intervals inevitably become wider with time. Consider survival beyond 2 years, that is $A(t) - A(2)$ for $t > 2$. The intervals would then have zero width at 2 years

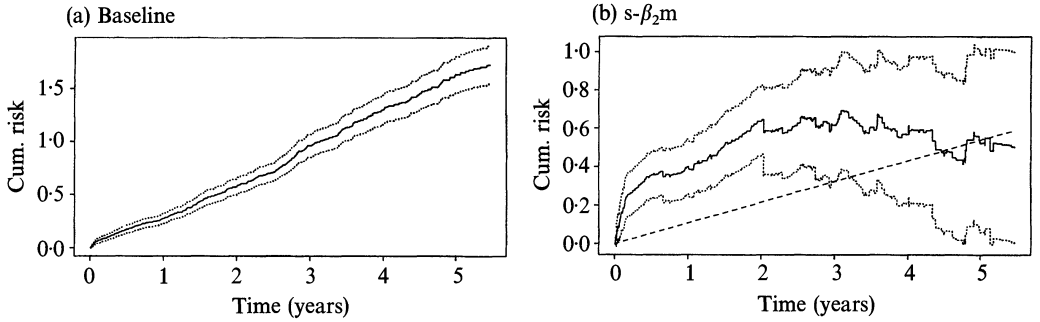

Fig. 1. Estimate of (a) the baseline cumulative risk and (b) the cumulative risk for s-β_2m, both shown by solid lines, with corresponding 95% pointwise confidence limits (dotted lines) based on the final model. The straight line estimate in (b) is obtained from the model in which Hb and s-β_2m are treated parametrically, and the baseline treated nonparametrically.

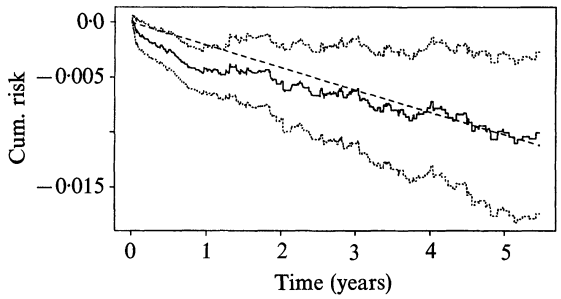

Fig. 2. Estimate of the cumulative risk for Hb (solid line) with corresponding 95% pointwise confidence limits (dotted lines) based on the 'full' Aalen model. The straight line estimate is obtained from the model with Hb treated parametrically, and s-β_2m and the baseline treated non-parametrically.

and would be narrower at 5 years. There would be a second set of bands, identical to the ones in Fig. 1(b), for 0 to 2 years. The two sets of intervals should be made wider to allow for the implicit multiple testing that is taking place:

(i) $A(t) = tb$ for $0 < t < 2$,

(ii) $A(t) - A(2) = (t-2)b$ for $2 < t < 6$.

To predict survival from our model, one can use the estimate

$$\hat{S}(t|x,z) = \exp\left\{-\int_0^t (x'\,d\hat{A} + z'\hat{\beta}\,ds)\right\}$$

of pr $(T > t|x,z)$ at given values of the covariates. Figure 3 shows the average predicted survival probabilities for groups defined in terms of the lower/upper quartiles of the covariates. Patients with low haemoglobin and high s-β_2m are at the highest risk.

Each curve in Fig. 3 is an average of estimated survival functions. They could be made monotone by isotonically regressing $\hat{S}(t|x,z)$ against t. We have not done that here, to show that the lack of monotonicity is only very slight. Indeed, an estimated survival curve

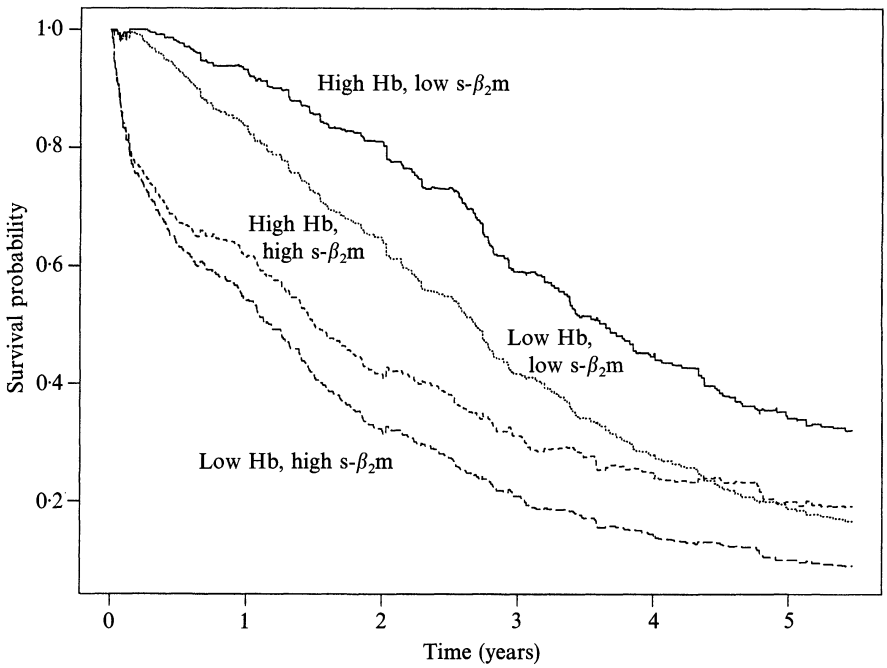

Fig. 3. Average predicted survival probabilities for four risk groups.

with significantly increasing sections would indicate a lack of fit, since on the model the estimate is consistent for the true survival function.

In Fig. 4 we plot the local Kaplan–Meier estimate for the high haemoglobin and low s-β_2m group (Hb > 122, s-β_2m < 3·3) and compare it with the average survival probabilities predicted by the different models as a rough check of goodness of fit. From this and similar plots for other groups, it appears that our model offers a better fit than either the Cox or the Lin–Ying model (1·2).

The estimated relative efficiencies of the weighted least squares estimators compared to the ordinary least squares estimators were 114% and 120% for the baseline and s-β_2m cumulative hazard functions at four years, and 112% for the parameter corresponding to haemoglobin.

We also analyzed data on 559 patients from the fifth trial with five additional covariates: indicators for treatment, sex and four age strata. These were included parametrically. The treatment, a trial drug regimen, was compared to conventional chemotherapy. The baseline and s-β_2m were handled nonparametrically. Haemoglobin was tried parametrically, but was not significant and was dropped. The shape of the s-β_2m cumulative hazard curve, in Fig. 5(a), is similar to that in the fourth trial, in Fig. 1(b): a plateau after about 2·5 years. The curve is plotted only up to 3·5 years due to shorter follow-up in the fifth trial. From Fig. 5(b) the treatment effect appears to be constant, so it is handled parametrically.

The Wald statistic for the treatment effect was −2·99, suggesting that patients on the drug regimen had significantly better survival (MacLennan et al., 1992). The predicted effect of treatment is to increase to probability of survival at two years by approximately 30%.

Our model gives a new interpretation to the data in keeping with scientific knowledge

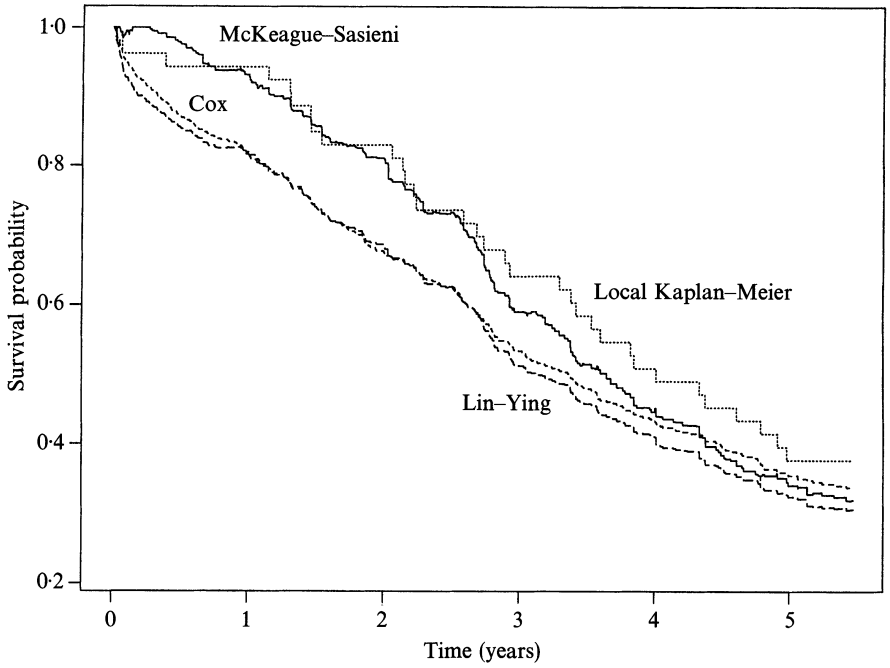

Fig. 4. Local Kaplan–Meier estimate of survival probability for the high Hb/low s-β_2m risk group, compared with various model-based estimates averaged over this group.

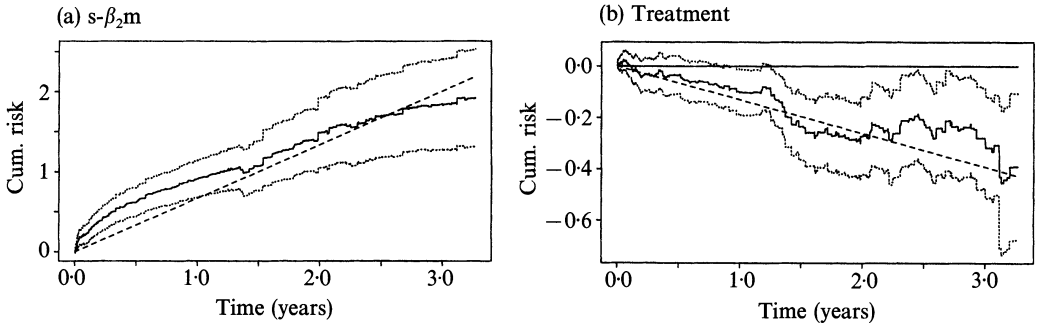

Fig. 5. Fifth myelomatosis trial based estimates of the cumulative risk (a) for s-β_2m, and (b) for the treatment effect. Compare Fig. 5(a) with Fig. 1(a).

of the course of disease. Although it is a model of convenience, and not based on some prior scientific theory, it yields insights that might have been missed by conventional analysis. A simple proportional hazards model does not adequately fit the data. A reasonable fit can be obtained by using a separate proportional hazards models for each year of survival (Cuzick et al., 1990), but such an approach is rather arbitrary and does not make efficient use of the data.

5. Discussion

The standard method for regression analysis of survival data is the proportional hazards model with exponential link (Cox, 1972). Comparisons with the present model are in order.

Consider first a single binary covariate representing two samples. The nonparametric additive model permits nonparametric estimation of the survival function in each sample separately. The Cox model permits a single nonparametric hazard function and assumes that the hazard in one sample is at all times a constant multiple of the hazard in the other. An additive model with a nonparametric baseline and parametric covariate effect is similar to the Cox model, except that the difference between the two hazard functions, rather than their ratio, is constant over time (Table 2).

Table 2. *Model assumptions for the two-sample problem*

Aalen:	λ_1, λ_2 unspecified
Cox:	$\lambda_2(t) = \theta\lambda_1(t)$, λ_1 unspecified
New:	$\lambda_2(t) = \lambda_1(t) + \theta$, λ_1 unspecified

The flexibility of our approach can be seen by comparing it to a stratified Cox model $\lambda(t|x, z) = \lambda(t|x)\exp(\beta' z)$, where x is stratum membership. In (1·1) we would take x to be the vector of indicators for membership in each stratum, so that each component of α represents a different stratum.

It is possible to generalize the Cox model so that it is directly comparable to (1·1). Consider

$$\lambda(t|x, z) = \lambda_0(t)\exp\{\alpha(t)'x + \beta' z\},$$

with unknown $\alpha(.)$ and β. This is a partly parametric version of a model studied by Zucker & Karr (1990). A histogram sieve approach can be used to fit this model: treat α as a step function, constant on each of K intervals $\mathscr{I}_j$ that partition the follow-up period, and compare with the grouped data version of our model. This gives a standard Cox model problem with $Kq + p$ covariates defined by the Kq components of the $x1_{\mathscr{I}_j}$ and the p components of z. Asymptotic theory, with K as well as n tending to infinity, can be developed along the lines of Murphy & Sen (1991).

Appendix 1

Asymptotic distributions

The asymptotic distributions of $\hat{\beta}$ and $\hat{A}$ are obtained under conditions stated by McKeague (1988a) or Huffer & McKeague (1991). In particular, assume that the covariates are bounded and $\lambda(.|x, z)$ is bounded away from zero. The follow-up period is taken to be a fixed finite interval. Let $M = (M_1, \ldots, M_n)'$, where

$$M_i(t) = N_i(t) - \int_0^t 1_{[T_i \geqslant s]}\lambda_i(s)\,ds$$

is the martingale associated with the counting process N_i.

We begin by noting that

$$\int Z'\hat{H}\,dM = \int Z'\hat{H}\,dN - \int Z'\hat{H}X\,dA - \int Z'\hat{H}Z\,dt\,\beta$$

$$= \int Z'\hat{H}\,dN - \int Z'\hat{H}Z\,dt\,\beta$$

since $\hat{H}$ is orthogonal to X. Hence

$$n^{\frac{1}{2}}(\hat{\beta}-\beta)=\left(n^{-1}\int Z'\hat{H}Z\,dt\right)^{-1}n^{-\frac{1}{2}}\int Z'\hat{H}\,dM, \tag{A1·1}$$

provided the inverse matrix exists. This will allow us to obtain the asymptotic distribution of $\hat{\beta}$ for any predictable $\hat{W}$ that is a uniformly consistent estimate of W, via the martingale central limit theorem. The weights computed via Method 1 are predictable. Those from Method 2 are not predictable since they depend on the initial estimate of β, both explicitly and through the initial estimate of α. An additional argument at the end of this appendix justifies replacement of the initial estimate $\tilde{\beta}$ by β.

Let $Y=(X, Z)$. As a consequence of the independent and identically distributed replicates, $n^{-1}Y'WY$ converges in probability to a nonrandom matrix function uniformly over bounded time intervals. This function is assumed to be nonsingular and smooth. The martingale central limit theorem can be applied to $n^{-\frac{1}{2}}\int Z'\hat{H}\,dM$, where the integral is over the range $(0,.)$, which has predictable variation

$$\left\langle n^{-\frac{1}{2}}\int_0^{\cdot} Z'\hat{H}\,dM\right\rangle_t = n^{-1}\int_0^t Z'\hat{H}W^{-1}\hat{H}'Z\,ds.$$

Routine matrix algebra gives $\hat{H}\hat{W}^{-1}\hat{H}'=\hat{H}$. Also,

$$n^{-1}Z'\hat{H}(\hat{W}^{-1}-W^{-1})\hat{H}'Z$$

converges uniformly in probability to zero (McKeague, 1988b, Lemma 4.3). Let Σ denote the limit in probability of $n^{-1}\int Z'HZ\,dt$. By uniform consistency of $\hat{H}$ and boundedness of the covariates, the matrix $n^{-1}\int Z'\hat{H}Z\,dt$ also converges in probability to Σ. It follows from (A1·1) that $n^{\frac{1}{2}}(\hat{\beta}-\beta)$ converges in distribution to a mean zero multivariate normal with variance Σ^{-1}.

From (2·3) and the definition of $\hat{A}$,

$$n^{\frac{1}{2}}(\hat{A}-A)=n^{\frac{1}{2}}\int_0^{\cdot}(X'\hat{W}X)^{-1}X'\hat{W}\,dM-\int_0^{\cdot}(X'\hat{W}X)^{-1}X'\hat{W}Z\,dt\,n^{\frac{1}{2}}(\hat{\beta}-\beta), \tag{A1·2}$$

provided the inverse matrix exists at all t; if not, an additional term of order $o_p(1)$ is required. Once again we can apply the martingale central limit theorem. The covariation between

$$n^{\frac{1}{2}}\int_0^{\cdot}(X'\hat{W}X)^{-1}X'\hat{W}\,dM$$

and

$$n^{-\frac{1}{2}}\int_0^{\cdot}Z\hat{H}\,dM$$

is

$$\int_0^{\cdot}(X'\hat{W}X)^{-1}X'\hat{W}W^{-1}\hat{H}'Z\,dt,$$

which converges in probability to a matrix of zeros, by the uniform consistency of $\hat{W}$ and the orthogonality of $\hat{H}$ and X. Thus the two terms on the right-hand side of (A1·2) are asymptotically independent. Let V denote the limit in probability of $n^{-1}X'WX$. The first term in (A1·2) is simply the limit of $n^{\frac{1}{2}}(\hat{A}-A)$ in the usual additive risk model in which $\beta=0$. It converges in distribution to a Gaussian martingale m with covariation process $\int V^{-1}\,dt$, where the integral is over $(0,.)$. It follows that $n^{\frac{1}{2}}(\hat{A}-A)$ converges in distribution to $m+\psi(.)\xi$, where m and ξ are independent, ξ is mean zero multivariate normal with variance Σ^{-1}, and $\psi(t)=\int V^{-1}U\,ds$ where the integral is over the range $(0, t)$, and where U is the limit in probability of $n^{-1}X'WZ$.

It remains to modify the above argument to allow for the nonpredictability of $\hat{W}$ when the weights are computed via Method 2. First consider the last part of (A1·1), $n^{-\frac{1}{2}}\int Z'\hat{H}\,dM$, each component of which can be written in the form

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\int G_i(\tilde{\beta})\,dM_i, \tag{A1·3}$$

where $G_i(\beta)=G_i(\beta,t)$ is predictable, twice differentiable in β, and $\tilde{\beta}$ is the initial estimate of β. Taylor expanding G_i about the true β, we can express (A1·3) as

$$n^{-\frac{1}{2}}\sum\int G_i(\beta)\,dM_i+n^{\frac{1}{2}}(\tilde{\beta}-\beta)'n^{-1}\sum\int\dot{G}_i(\beta)\,dM_i$$

$$+n^{-\frac{1}{2}}\{n^{\frac{1}{2}}(\tilde{\beta}-\beta)\}'\left\{n^{-1}\sum\int\ddot{G}_i(\beta^*)\,dM_i\right\}\{n^{\frac{1}{2}}(\tilde{\beta}-\beta)\},$$

where β^* lies on the line segment between β and $\tilde{\beta}$, and the dependence of β^* on t and i has been suppressed. The first term, having predictable integrands, can be treated using the martingale central limit theorem as before. The second term is easily shown to converge to zero in probability since $\tilde{\beta}$ is $n^{\frac{1}{2}}$-consistent and the integrand is predictable and uniformly bounded. The third term is also asymptotically negligible since $\ddot{G}_i$ is uniformly bounded in a neighbourhood of β. A similar argument applies to the first term on the right-hand side of (A1·2). We conclude that the asymptotic distribution of $\hat{\beta}$ and $\hat{A}$ is the same for Methods 1 and 2.

Appendix 2

The efficient score for β

The efficient score for β is obtained by projecting the score $\dot{l}_\beta$ onto the orthogonal complement of the tangent space spanned by the range of the score operator $\dot{l}_\alpha$. When $n=1$, it will be given by

$$l_\beta^*=\int\frac{z-b^*(t)'x}{\lambda(t|x,z)}\,dM(t|x,z)$$

for some b^* such that l_β^* is orthogonal to $\dot{l}_\alpha b$ for all b such that

$$E[\delta\{b(T)'x/\lambda(T|x,z)\}^2]<\infty.$$

That is, for all such b,

$$0=E\left\{\int\frac{z-b^*(t)'x}{\lambda(t|x,z)}\,dM(t)\int\frac{b(t)'x}{\lambda(t|x,z)}\,dM(t)\right\}$$

$$=E\left\{\delta\frac{(z-b^*(T)'x)}{\lambda(T|x,z)}\frac{b(T)'x}{\lambda(T|x,z)}\right\};$$

see, for example, Sasieni (1992, Lemma A.1). Hence

$$b^*(t)'=E\left\{\frac{zx'}{\lambda^2(t|x,z)}\,\middle|\,T=t,\delta=1\right\}\left[E\left\{\frac{xx'}{\lambda^2(t|x,z)}\,\middle|\,T=t,\delta=1\right\}\right]^{-1}$$

$$=E\left\{\frac{zx'}{\lambda(t|x,z)}1_{[T\geqslant t]}\right\}\left[E\left\{\frac{xx'}{\lambda(t|x,z)}1_{[T\geqslant t]}\right\}\right]^{-1}$$

(Sasieni, 1992, § 3). The variance bound $(El_\beta^*l_\beta^{*\prime})^{-1}$ is attained by the asymptotic variance of $\hat{\beta}$.

For a sample of size n, the efficient score for β is approximately

$$\hat{l}^*_\beta = \int Z'W\,dN - \int (Z'WX)(X'WX)^{-1}X'W\,dN - \int Z'WX\,dA$$
$$+ \int (Z'WX)(X'WX)^{-1}X'WX\,dA - \int Z'WZ\,dt\,\beta + \int (Z'WX)(X'WX)^{-1}X'WZ\,dt\,\beta$$
$$= \int Z'H\,dN - \int Z'HZ\,dt\,\beta.$$

Solving $\hat{l}^*_\beta = 0$ for β gives (2·4). Thus the estimators $\hat{\beta}$ and $\hat{A}$ discussed in this paper are asymptotically efficient for the semiparametric model (Bickel et al., 1993).

REFERENCES

AALEN, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory*, Lecture Notes in Statistics, **2**, Ed. W. Klonecki, A. Kozek and J. Rosinski, pp. 1–25. New York: Springer-Verlag.

AALEN O. O. (1989). A linear regression model for the analysis for life times. *Statist. Med.* **8**, 907–25.

AALEN O. O. (1993). Further results on the nonparametric linear regression model in survival analysis. *Statist. Med.* **12**, 1569–88.

BARLOW, R. E. & PRENTICE, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75**, 65–74.

BICKEL, P. J., KLAASEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models.* Baltimore: Johns Hopkins University Press.

COOK, R. D. & WEISBERG, S. (1982). *Residuals and Influence in Regression.* London: Chapman & Hall.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–220.

CUZICK, J., DE STAVOLA, B. L., COOPER, E. H., CHAPMAN, C. & MACLENNAN, I. C. M. (1990). Long-term prognostic value of serum β_2 microglobulin in myelomatosis. *Br. J. Haemat.* **75**, 506–10.

GREENWOOD, P. E. & WEFELMEYER, W. (1990). Efficiency of estimators for partially specified filtered models. *Stoch. Process Applic.* **36**, 353–70.

GREENWOOD, P. E. & WEFELMEYER, W. (1991). Efficient estimating equations for nonparametric filtered models. In *Statistical Inference in Stochastic Processes*, **1**, Ed. N. U. Prabhu and I. V. Basawa, pp. 107–41. New York: Marcel Dekker.

HENDERSON, R. & MILNER, A. (1991). Aalen plots under proportional hazards. *Appl. Statist.* **40**, 401–10.

HENDERSON, R. & OMAN, P. (1993). Influence in linear hazard models. *Scand. J. Statist.* **20**, 195–212.

HUFFER, F. W. & MCKEAGUE, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Am. Statist. Assoc.* **86**, 38–53.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

LIN, D. Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

MAU, J. (1986). On a graphical method for the detection of time-dependent effects of covariates in survival data. *Appl. Statist.* **35**, 245–55.

MAU, J. (1988). A comparison of counting process models for complicated life histories. *Appl. Stoch. Models Data Anal.* **4**, 283–98.

MACLENNAN, I. C. M., CHAPMAN, J., DUNN, J. & KELLY, K. (1992). Combined chemotherapy with ABCM versus melphalan for treatment of myelomatosis. *Lancet* **339**, 200–5.

MCKEAGUE, I. W. (1988a). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *Contemp. Math.* **80**, 139–52.

MCKEAGUE, I. W. (1988b). A counting process approach to the regression analysis of grouped survival data. *Stoch. Process Applic.* **28**, 221–39.

MEDICAL RESEARCH COUNCIL (1984). Analysis and management of renal failure in the fourth MRC myelomatosis trial. *Br. Med. J.* **288**, 1411–6.

MURPHY, S. A. & SEN, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stoch. Processes Applic.* **39**, 153–80.

SASIENI, P. D. (1992). Information bounds for the additive and multiplicative intensity models. In *Survival Analysis: State of the Art*, Ed. J. P. Klein and P. K. Goel, pp. 249–65. Dordrecht, Boston, London: Kluwer.

THERNEAU, T. M., GRAMBSCH, P. M. & FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–60.

ZUCKER, D. M. & KARR, A. F. (1990). Survival data regression analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18**, 329–53.

[*Received February* 1993. *Revised January* 1994]