REPORT DOCUMENTATION PAGE

(2)

Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) FSU Technical Report M-882	5. MONITORING ORGANIZATION REPORT NUMBER(S) DAAL03-90-G-0103
6a. NAME OF PERFORMING ORGANIZATION Florida State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics Tallahassee, FL 32306-3033	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U. S. Army Research Office 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER ARO 27868.23-MA
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
A PARTLY PARAMETRIC ADDITIVE RISK MODEL

12. PERSONAL AUTHOR(S)
Ian W. McKeague, Peter D. Sasieni

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	January 1993	24

16. SUPPLEMENTARY NOTATION
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Aalen's linear hazards model, counting processes, right-censored data, semiparametric.

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

ABSTRACT. Aalen's additive risk model allows the influence of covariates on a hazard function to vary over time, and to do so in a different fashion for each covariate. Although allowing greater flexibility than a Cox model, which has a more parsimonious temporal structure, the number of covariates that can be handled by Aalen's model is quite limited. One way around this difficulty is to impose some *a priori* structure on the form of the model, thereby reducing the number of functions to be estimated. In this paper we introduce a partly parametric version of Aalen's model in which only a small number of the covariates are selected to have their influence vary nonparametrically over time, and the influence of the remaining covariates is restricted to be constant in time. Efficient procedures for fitting this new model are developed and studied. The approach is applied to data from the British Medical Research Council's myelomatosis trials.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Ian W. McKeague	22b. TELEPHONE (Include Area Code) 904-644-3218 / 22c. OFFICE SYMBOL

A PARTLY PARAMETRIC ADDITIVE RISK MODEL

Ian W. McKeague [1]
Department of Statistics
Florida State University
Tallahassee, FL 32306-3033

Peter D. Sasieni
Imperial Cancer Research Fund
Lincoln's Inn Fields
London WC2A 3PX, England

FSU Technical Report No. M-882
USARO Technical Report No. D-131
AFOSR Technical Report No. 91-277

January, 1993

[1]Research partially supported by Army Research Office Grant DAA03-90-G0103 and by the Air Force Office of Scientific Research under Grant AFOSR91-0048.

A PARTLY PARAMETRIC ADDITIVE RISK MODEL

IAN W. MCKEAGUE* AND PETER D. SASIENI **

**Florida State University, Tallahassee* ***Imperial Cancer Research Fund, London*

ABSTRACT. Aalen's additive risk model allows the influence of covariates on a hazard function to vary over time, and to do so in a different fashion for each covariate. Although allowing greater flexibility than a Cox model, which has a more parsimonious temporal structure, the number of covariates that can be handled by Aalen's model is quite limited. One way around this difficulty is to impose some *a priori* structure on the form of the model, thereby reducing the number of functions to be estimated. In this paper we introduce a partly parametric version of Aalen's model in which only a small number of the covariates are selected to have their influence vary nonparametrically over time, and the influence of the remaining covariates is restricted to be constant in time. Efficient procedures for fitting this new model are developed and studied. The approach is applied to data from the British Medical Research Council's myelomatosis trials.

MSC 1991 subject classifications. Primary: 62G05; Secondary: 62M09.

Key words and phrases. Aalen's linear hazards model, counting processes, right-censored data, semiparametric.

1 Introduction

Aalen (1980) proposed the following additive model for the intensity of a counting process:

$$\lambda(t|z) = \alpha(t)'z,$$

where z is a p-vector of covariates and α is a p-vector of unknown functions of time. The first component of z may be set to 1 to allow for a baseline hazard. More recently estimation in this model has been studied by Huffer and McKeague (1991) and McKeague (1988a, 1988b). Greenwood and 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 model has had only limited use in data analysis and primarily in data sets with just a few covariates. Examples may be found in Aalen (1989), Mau (1986, 1988) and Henderson and Milner (1991). One reason for the lack of use is that a separate nonparametric function must be estimated in association with each covariate. As a way of redressing this problem we here introduce a partly parametric version of the additive risk model in which the effects of some covariates are assumed to be constant in time. This restriction could be relaxed at some price by allowing a parametric model for a component $\alpha_i(\cdot)$. Alternatively, by defining a new time dependent covariate, $z^*(t) = z\exp(-t)$ say, non-constant time effects could be fit with this model directly.

Assume that the intensity at t for an individual with covariates x and z is given by

$$\lambda(t|x,z) = \alpha(t)'x + \beta'z, \tag{1.1}$$

with the covariates being q and p dimensional respectively. We are interested in estimating β and the vector of 'cumulative hazards' $A(\cdot) = \int_0^{\cdot} \alpha(s)\,ds$. If β were known then one could use Aalen's least squares estimator for $A(\cdot)$ followed by the Huffer–McKeague scheme to obtain an efficient estimator. Similarly, if $\alpha(\cdot)$ were known then one could estimate β by maximum likelihood. However, neither A nor β are known and it is not obvious how to construct efficient estimates although, intuitively, iterating between estimation of β and α should work. The approach used here is to look at the efficient score equation for β and to use this to obtain a set of pseudo-normal equations. There are similarities with the approach used by Sasieni (1992) to motivate the Huffer–McKeague estimator.

Various authors, most recently Lin and Ying (1992), have considered the following additive analogue of Cox's (1972) proportional hazards model:

$$\lambda(t|z) = \alpha_0(t) + \beta' z. \tag{1.2}$$

This is a special case of the partly parametric additive risk model (1.1) obtained by treating the baseline hazard function nonparametrically and all the covariates parametrically. The temporal influence of each covariates is required to be constant, so it is considerably less versatile than (1.1). However, this model, like the Cox model, can be useful in the initial exploratory stages when there are a very large number of covariates. In an application to real data (Section 4) we used (1.2) to find the two most influential covariates, followed by (1.1) with one of the covariates treated nonparametrically. This led to a much better fit than could be achieved by either (1.2) or the Cox model.

The paper is organised as follows. In Section 2 we derive our efficient estimators for β and A and discuss their practical implementation. In Section 3 we discuss ways to refine a partly additive risk model and a diagnostic technique which can give useful information about how much the estimates would change if an individual observation were removed from the data set. An application to finding prognostic factors for survival among myelomatosis patients is discussed in Section 4. Some general discussion, comparing the new approach with the standard Cox model approach to regression analysis of censored survival data, is provided in Section 5. The asymptotic distributions of the estimators are obtained in Section 6.

2 Semiparametric estimators

2.1 Notation and derivation of the 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 independent and identically distributed individuals: $\delta_i = 1$ if T_i is uncensored. In the usual survival set-up with non-informative and conditionally independent censoring the log-likelihood for λ is

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

where the range of integration extends over the period of follow-up and $\lambda_i(t) = \lambda(t|x_i, z_i)$. Let $N_i(t) = 1_{[T_i \le t, \delta_i = 1]}$ be the corresponding counting process and

$M_i(t) = N_i(t) - \int_0^t 1_{[T_i \geq s]} \lambda_i(s)\, ds$ the associated martingale. Assume that $\lambda(\cdot|x,z)$ is bounded away from zero. Consider a one-dimensional parametric submodel, $\alpha(t) = \alpha(t;\eta)$, in which

$$\frac{\partial \alpha(t)}{\partial \eta} = b(t).$$

Differentiating the log likelihood with respect to β and η we obtain the parametric score function:

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

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

$$\beta = (\int Z'WZ\, dt)^{-1} (\int Z'W\, dN - \int Z'WX\, dA), \tag{2.1}$$

where $Z = Z(t) = (z_1 1_{[T_1 \geq t]}, \ldots, z_n 1_{[T_n \geq t]})'$, the $n \times n$-matrix W is defined by $W(t) = \mathrm{diag}\{1/\lambda_i(t)\}$, $N(t) = (N_1(t), \ldots, N_n(t))'$ and X is defined like Z. Next,

$$\begin{aligned} \dot{l}_\eta = \dot{l}_\alpha b &= \int b(t)' X'W\, dM(t) \\ &= \int b'X'W\, dN - \int b'X'WZ\beta\, dt - \int b'X'WX\, dA, \end{aligned}$$

where $M = (M_1, \ldots, M_n)'$.

Setting $\dot{l}_\alpha b = 0$ for a sufficiently large collection of vector-valued functions b implies that

$$A(t) = \int_0^t (X'WX)^{-1} (X'W\, dN - X'WZ\beta\, ds). \tag{2.2}$$

Substituting the right hand side of (2.2) into (2.1) and solving for β gives

$$\hat{\beta} = (\int Z'HZ\, dt)^{-1} \int Z'H\, dN, \tag{2.3}$$

where $H = W - WX(X'WX)^{-1}X'W$. Now $\hat{\beta}$ is not an estimator since it depends on λ, which is unknown. However $\hat{\beta}$ is the solution of $l_\beta^* = 0$, where l_β^* is the efficient

score for β, as shown in the Appendix. This implies that an estimator based on (2.3), but with a consistent estimate of λ replacing the unknown function, will be efficient for the semiparametric model; see Bickel, Klaassen, Ritov and Wellner (1992, Theorem 3.4.1).

The use of the identity matrix I in place of W gives an initial estimate of β. To construct an efficient estimator we propose two methods. These both replace W by $\widehat{W} = \text{diag}\{1/\hat{\lambda}_i(\cdot)\}$ where $\hat{\lambda}_i$ is some estimate of λ_i. The second method is more appropriate when the dimension of β is large.

Method I:

(i) Fit the full Aalen model, $\lambda(t|x,z) = \alpha(t)'x + \beta(t)'z$, and obtain an estimate $\widehat{W}$ of W from a ***historical*** kernel smoother, as in Huffer and McKeague (1991).

(ii) Find an estimate $\hat{\beta}$ of β by (2.3), using $\widehat{W}$ in place of W.

(iii) Estimate A from (2.2) using $\widehat{W}$ and $\hat{\beta}$ in place of W and β.

Method II:

(i) Estimate β inefficiently from (2.3) using I in place of W.

(ii) Estimate A inefficiently from (2.2) using I in place of W and the estimate of β from (i), and then use historical kernel smoothing to estimate α.

(iii) Obtain an estimate $\widehat{W}$ of W using the estimates of β and α from (i) and (ii).

(iv) Obtain final estimates $\hat{A}$ and $\hat{\beta}$ using (2.2) and (2.3) with $\widehat{W}$ in place of W.

In our computer implementation we have used method II with a $\hat{\lambda}_i$ that is explicitly defined in subsection 2.4. Notice that Z and X are functions of t and that, provided the covariates are predictable, the same estimating equations (2.2) and (2.3) could be used with time-dependent covariates.

The gain in efficiency of the two-step estimator using $\widehat{W}$ compared to the initial estimator using I will depend on the heterogeneity of the hazard of individuals in the sample. For instance, if all individuals are at equal risk, so that none of the covariates are related to survival, then there is no efficiency gain. In general, however, there will be a small gain. Huffer and McKeague (1991) investigated by simulation the asymptotic relative efficiency of the OLS 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 somewhat more complicated here because our estimate of β depends on the weights for all individuals at risk at each failure time. In any given situation one can however

fairly easily examine the efficiency gain by comparing the asymptotic covariance matrices or more thoroughly via a bootstrap simulation.

2.2 Estimating the asymptotic covariance matrix

The asymptotic distribution of $\hat{\beta}$ and $\hat{A}$ is the same for methods I and II. Standard counting process techniques (Section 6) can be used to show that $n^{1/2}(\hat{\beta} - \beta)$ converges in distribution to a p-variate normal with mean zero and with covariance matrix which can be consistently estimated by $\widehat{\Sigma}^{-1}$, where $\widehat{\Sigma} = n^{-1} \int Z' \widehat{H} Z \, dt$. Here $\widehat{H}$ is the estimated version of H obtained by replacing W by a consistent estimate $\widehat{W}$. The same approach shows that $n^{1/2}(\hat{A} - A)$ converges in distribution to a q-variate Gaussian process with mean zero and with a covariance function which, as a function of s and t, can be consistently estimated by

$$n \sum_{u \leq s \wedge t} \Delta_u \Delta_u' + \hat{\psi}(s) \widehat{\Sigma}^{-1} \hat{\psi}(t)', \tag{2.4}$$

where Δ_u is the jump in $\hat{A}$ at time u and

$$\hat{\psi}(t) = \int_0^t (X' \widehat{W} X)^{-1} X' \widehat{W} Z \, ds.$$

The first term in (2.4) is a consistent estimate of the covariance function for the model in which only the nonparametric terms are non-zero; the second term represents the contribution from the parametric part of the model.

2.3 Grouped data version

Although evaluation of the estimates is fast on even a small computer, one may wish to fit a grouped data version of the model in the exploratory stage of model building. For most purposes grouping the time axis into about ten intervals will be adequate and this will greatly reduce the computation. The grouped data model may be written

$$\lambda(t|x,z) = \sum_{i=1}^{K} \alpha_{(i)}' x 1_{\mathcal{I}_i}(t) + \beta' z,$$

where the time axis has been divided into K intervals $\mathcal{I}_1, \ldots, \mathcal{I}_K$ with $\mathcal{I}_i = [\tau_{i-1}, \tau_i)$ and $\tau_0 = 0$. One approach to estimation treats this as a parametric linear model with $Kq + p$ parameters, but even for moderately large K it makes sense to take

into account the orthogonality of the dummy covariate blocks $x1_{\mathcal{I}_i}(t)$, $i = 1, \ldots, K$. Let $k(u)$ denote the index such that $u \in \mathcal{I}_{k(u)}$. Proceeding as before one has

$$A(u) = \sum_{i=1}^{k(u)-1} \alpha_{(i)}\tau_i + \alpha_{k(u)}(u - \tau_{\{k(u)-1\}})$$

and

$$\alpha_{(i)} = \Big(\int_{\mathcal{I}_i} X'WX\, dt\Big)^{-1} \int_{\mathcal{I}_i} (X'W dN - X'WZ\beta\, du).$$

Thus instead of having to solve a system of $(p+q)$ linear equations at each failure time, one has only to solve such a system for each time interval.

2.4 The choice of weights

Although asymptotically one may use any consistent $\hat{\lambda}_i$ obtained via method I or II to estimate the efficient weights, in practice the choice of $\hat{\lambda}_i$ needs to be made with some care. It is a good idea to compare the weighted estimates with the unweighted ones, since both are consistent on the model, and this can provide a check of whether the weights are wildly off-target.

The implementation that we have used calculates $\hat{\lambda}_i$ as follows. Let $T_{(i)}$ denote the ith ordered failure time, and set $T_{(0)} = 0$. Given initial estimates $\tilde{A}$ and $\tilde{\beta}$, we estimate $\alpha(t)$ for $t > T_{(d)}$ by

$$\tilde{\alpha}(t) = \frac{\hat{A}(T_{(i)}) - \hat{A}(T_{(i-d)})}{T_{(i)} - T_{(i-d)}} \quad \text{when } T_{(i)} < t \le T_{(i+1)}.$$

We have found that taking d between $n^{1/2}$ and $4n^{1/2}$ 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 cannot always be used to estimate the weights since it may be non-positive and it is undefined for $t \le T_{(d)}$. Instead, we use

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

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 defining the 'minimum allowable hazard' $\epsilon\bar{\lambda}(t)$. The examples in this paper are based on $d = 50$ and $\epsilon = 0.25$. Strictly speaking, $\hat{\lambda}_i$ departs from methods I and II for $t \le T_{(d)}$, but this will have negligible effect provided that $T_{(d)}$ is small compared to the total length of follow-up.

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 failure times.

3 Model refinement and diagnostics

The most immediate problem faced when applying the partly parametric additive risk model is in determining which of the covariates should be modeled nonparametrically. There may be scientific reasons for wanting to include a particular variable nonparametrically. It is possible that some factor will not be significant when modeled 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. However, if there are a large number of covariates, then it is advisable to start by treating at most a few of them nonparametrically and the rest parametrically. It is generally sensible to include a nonparametric baseline. The covariates having the most insignificant effects should then be dropped from the model one-by-one. The next step would be to examine whether the influence of each of the parametric covariates varies with time by treating each nonparametrically and looking 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 against time will give some indication of the validity of the parametric assumptions and how they are violated when they fail. This approach is illustrated by the example in the following section. Other approaches are possible. For instance, one might fit a separate Aalen model for each covariate, to get an initial idea of the variation of the additive 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 can give useful information about how much the estimates would change if an individual observation were removed from the data set. For fixed W, $\hat{\beta}$ as defined in (2.3) is an explicitly defined functional of the empirical distribution function. Differentiating this functional and evaluating the derivative at the empirical distribution gives the empirical influence curve (Cook and Weisberg, 1982, pp. 104–108). Straightforward differentiation and a little algebra yields

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

as the influence of the ith individual on $\hat{\beta}$. Here $\hat{M}_i(t) = N_i(t) - \int_0^t 1_{[T_i \geq s]}(x_i' \, d\hat{A} + \hat{\beta}' z_i \, ds)$ is the martingale residual for observation i. 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 with covariates x and z then the above expression for the $\partial\hat{\beta}_i$'s will be correct to first order.

As in proportional hazards regression, the basic diagnostic building block is the counting process martingale residual $\hat{M}_i(\cdot)$ (Barlow and Prentice, 1988; Therneau, Grambsch and Fleming, 1990). A plot of the $\sum \hat{M}_i(t)$ against t can be used to check for systematic 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 one may partition the time axis into about ten intervals $[\tau_{j-1}, \tau_j)$ $(j = 1, \ldots, J)$ and for each j, 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 plots are analogous to partial residual plots in the linear model and may detect the need to transform a covariate. One can also check whether the additive risk associated with a given covariate varies in time by comparing the J plots for that covariate, after rescaling each by $\tau_j - \tau_{j-1}$.

4 Example

In this section we discuss the fitting of a partly parametric additive risk model to data from the British Medical Research Council's (1984) fourth myelomatosis trial. We analyzed survival data on 495 myelomatosis patients for whom presentation measurements included serum β_2 microglobulin and haemoglobin. Percentiles of these measurements are given in Table 1. In fitting the regression models, serum β_2 microglobulin was transformed by $\log_{10}(\cdot)$ to compensate for its skewness.

Several studies (e.g. Cuzick, Cooper and MacLennan, 1985) have indicated that serum β_2 microglobulin is of primary importance in predicting survival in myelomatosis patients. However, a recent paper of Cuzick, De Stavola, Cooper, Chapman and MacLennan (1990) suggests 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. We think that it is more appropriate to apply a partly parametric additive risk model when searching for such variations.

Table 1. Percentiles of serum β_2 microglobulin and haemoglobin.

Covariate	min	10	25	50	75	90	max
serum β_2	0.3	2.3	3.3	5.7	9	22	76.7
haemoglobin	25	71	90	106	122	136	167

We initially treated the covariates parametrically and the baseline nonparametrically, as in model (1.2). The Wald statistics for testing whether the corresponding parameters are zero were 2.25 for serum β_2 microglobulin and -3.24 for haemoglobin. Thus there is strong evidence that both covariates are influential.

Next we considered the model with haemoglobin treated parametrically, and the baseline hazard and serum β_2 microglobulin treated nonparametrically. This turned out to be our final model. Figures 1 and 2 show plots of the cumulative risks for the two nonparametric terms; Figure 2 also contains the straight line estimate of the cumulative risk for serum β_2 microglobulin based on model (1.2). Note that in the first three years the straight line falls outside the 95% confidence limits, strongly suggesting that the influence of serum β_2 microglobulin varies with time. Further inspection of Figure 2 indicates a plateau in the nonparametric cumulative risk estimate after about two years, which is consistent with the claim of Cuzick et al. (1990) that serum β_2 microglobulin is of primary importance in predicting survival only within the first two years of follow-up. Haemoglobin treated parametrically has significant influence (the Wald statistic is -3.13) and from Figure 3 we see that its influence does not vary appreciably with time since the straight line estimate of cumulative risk is almost completely contained in the 95% confidence limits around the nonparametric estimate based on the full Aalen model.

It is noticeable from Figure 2 that the confidence intervals inevitably become wider with time. Suppose one looked at survival beyond 2 years: i.e., $A(t) - A(2)$ for $t > 2$. In that way the intervals would have zero width at 2 years and would be narrower at 5 years. There would be a second set of bands, identical to the ones in Figure 2, for 0 to 2 years. It would seem sensible that 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 $0 < t < 2$ and (ii) $A(t) - A(2) = (t-2)b$ for $2 < t < 6$.

For the purpose of predicting survival based on our final model, one can use the estimate

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

of the survival function $S(t|x,z) = \mathrm{pr}(T > t|x,z)$ at given values of the covariates. In Figure 4 we have plotted the average predicted survival probabilities for groups defined in terms of the lower/upper quartiles of the covariates. Note that patients with low haemoglobin and high serum β_2 microglobulin are at the highest risk, whereas patients with high haemoglobin and low serum β_2 microglobulin are at the lowest risk.

A less than pleasing feature of the curves in Figure 4 is that they are not monotone. Since each curve is an averaged estimates of survival functions it would be sensible to take a monotone version as the final curve. This could easily be achieved by isotonically regressing $\hat{S}(t|x,z)$ against t. We have not chosen to do that here in order to show that the lack of monotonicity is only very slight. Indeed an estimated survival curve with significantly increasing sections would indicate a lack of fit, since it is known that on the model the estimate is consistent for the true survival function.

As a rough check of goodness-of-fit it is useful to compare the various model based estimates of survival probability with the local Kaplan–Meier estimate. In Figure 5 we plot the local Kaplan–Meier estimate for the high haemoglobin and low serum β_2 microglobulin group (Hb > 122, serum $\beta_2 < 3.3$) and compare it with the average survival probabilities predicted by the different models. It appears that our model offers a much better fit than either the Cox model or the Lin–Ying model.

In this example the relative efficiencies of our estimators compared to the OLS estimators were 114% and 120% for the baseline and serum β_2 cumulative hazard functions at four years, and 112% for the parameter corresponding to haemoglobin.

[Insert Figures 1–5 about here]

Figure 1. Estimate of the baseline cumulative risk (————) with 95% confidence limits (--------) based on the final model.

Figure 2. Estimate of the cumulative risk for serum β_2 (————) with corresponding 95% confidence limits (--------) based on the final model. The straight line

estimate is obtained from the model in which serum β_2, as well as Hb, are treated parametrically .

Figure 3. Estimate of the cumulative risk for Hb (————) with corresponding 95% confidence limits (-------) based on the "full" Aalen model. The straight line estimate is obtained from the model with Hb treated parametrically and serum β_2 and the baseline treated nonparametrically.

Figure 4. Average predicted survival probabilities according to risk group.

Figure 5. Local Kaplan–Meier estimate of survival probability for the high Hb/low serum β_2 group, compared with various model-based cstimates averaged over this group.

We have also applied our approach to data on 559 patients from the British Medical Research Council's fifth myelomatosis trial. In this case we used indicators for treatment, sex and four age strata as covariates entering parametrically. The treatment was a trial drug regimen that was compared to conventional chemotherapy. The baseline and serum β_2 were handled nonparametrically as before. Haemoglobin was tried parametrically, but did not turn out to be a significant covariate, so was dropped from the model. The shape of the serum β_2 cumulative hazard curve, given in Figure 6, is remarkably similar to that in the fourth trial (Figure 2): the straight line estimate again falls outside the 95% confidence intervals and there appears to be a plateau after about 2.5 years. The curve is plotted only up to 3.5 years, which is as far as we can go with data from the fifth trial. Figure 7 justifies our handling of the treatment parametrically: the treatment effect appears to be constant in time since the straight line falls within the 95% confidence intervals for the nonparametric cumulative hazard.

[Insert Figures 6 and 7 about here]

Figure 6. Fifth myelomatosis trial based estimates of the cumulative risk for serum β_2; compare with Figure 1.

Figure 7. Fifth myelomatosis trial based estimates of the cumulative risk for the treatment effect.

The Wald statistic for testing for a treatment effect was -2.99, suggesting that patients receiving the drug regimen had significantly better survival. At two years, the predicted effect of treatment is to increase the probability of survival by approximately 30% over what it would be for an individual on conventional chemotherapy.

5 Discussion

The standard method for regression analysis of survival data at present is the proportional hazards model with exponential link function (Cox, 1972). Some comparisons between this and the present model seem in order.

Consider first the simplest case of 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 baseline hazard function and assumes that the hazard, at any time t, in one sample is always a common multiple of the hazard, at the same time t, in the other sample. 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 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

More generally, comparison between the Cox and the partly parametric additive model is simplest when the only nonparametric component is the baseline. In that case the Cox model has $\lambda(t|z) = \lambda_0(t)\exp(\beta' z)$ and the additive model has $\lambda(t|z) = \lambda_0(t) + \beta' z$. Such models have been considered before (e.g., Cox and Oakes, 1984, p.74).

The full flexibility of the semiparametric additive model is seen by comparing it to the stratified Cox model. Given two one-dimensional of covariates, x and z, one may describe the stratified Cox model by

$$\lambda(t|x,z) = \lambda(t|x)\exp(\beta' z)$$

$$\lambda(t|x=k) = \lambda_k(t) \quad \text{assuming } x \in \{1,2,\ldots,K\}$$

and the additive model by

$$\begin{aligned}\lambda(t|x,z) &= \lambda(t|x) + \beta' z \\ \lambda(t|x=k) &= \alpha(t)'x \\ &= \lambda_k(t), \quad \text{the } k\text{th component of } \alpha(t) \\ &\qquad \text{when } x \text{ is a vector of } K \text{ dummy variables.}\end{aligned}$$

It is possible to generalise the Cox model so that it is directly comparable to our partly parametric Aalen model. Consider the Cox-type model

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

with time-dependent coefficients $\alpha(t)$ and time-independent coefficients β. This is a partly parametric version of a model studied by Zucker and Karr (1990). Note that $\exp\{\alpha(t)'(x_2 - x_1)\}$ can be interpreted as the time-specific relative risk between an individual with covariates (x_2, z) and one with (x_1, z). The unknown α and β can be estimated via a histogram sieve approach: treat α as a step function, constant on each of K intervals $\mathcal{I}_i$ that partition the follow-up period, cf. 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 $x1_{\mathcal{I}_i}$ and the p components of z. Asymptotic theory, with K as well as n tending to infinity, for the resulting sieve estimators can be developed along the lines of Murphy and Sen (1991), who studied the fully time-dependent case. We shall not pursue this here, but we note that the asymptotic theory is considerably more complicated to develop than with the partly parametric additive risk model.

6 Asymptotic distributions

In this section we find the asymptotic distribution of our estimators $\hat{\beta}$ and $\hat{A}$. This is done under conditions stated in McKeague (1988a) or Huffer and McKeague (1991); in particular, the covariates are assumed to be bounded and $\lambda(\cdot|x,z)$ is assumed to be bounded away from zero. The follow-up period is taken to be a fixed bounded interval.

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

$$n^{1/2}(\hat{\beta} - \beta) = \left(n^{-1} \int Z'\hat{H} Z\, dt\right)^{-1} n^{-1/2} \int Z'\hat{H}\, dM, \tag{5.1}$$

provided the inverse matrix exists. This will allow us to obtain the asymptotic distribution of $\hat{\beta}$ for any predictable $\widehat{W}$ that is a uniformly consistent estimate of W, via the martingale central limit theorem. For now suppose that the weights are computed via method I, in which case $\widehat{W}$ *is* predictable. The method II weights are not predictable because they are a function of the initial estimate of β, both explicitly and through dependence on the initial estimate of α, so some additional work will be needed in that case.

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. We apply the martingale central limit theorem to the martingale $n^{-1/2} \int_0^{\cdot} Z'\hat{H}\, dM$, which has predictable variation

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

Routine matrix algebra gives $\hat{H}\widehat{W}^{-1}\hat{H}' = \hat{H}$. Also, $n^{-1}Z'\hat{H}(\widehat{W}^{-1} - W^{-1})\hat{H}Z$ converges uniformly in probability to zero, cf. 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 (5.1) that $n^{1/2}(\hat{\beta} - \beta)$ converges in distribution to a mean zero multivariate normal with variance Σ^{-1}.

From (2.2) and the definition of $\hat{A}$,

$$\begin{aligned} n^{1/2}(\hat{A} - A) = n^{1/2} \int_0^{\cdot} (X'\widehat{W}X)^{-1} X'\widehat{W}\, dM \\ - \int_0^{\cdot} (X'\widehat{W}X)^{-1} X'\widehat{W} Z\, dt\, n^{1/2}(\hat{\beta} - \beta), \end{aligned} \tag{5.2}$$

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 $n^{1/2}\int_0^{\cdot}(X'\widehat{W}X)^{-1}X'\widehat{W}\,dM$ and $n^{-1/2}\int_0^{\cdot} Z\hat{H}\,dM$ is

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

which converges in probability to a matrix of zeros, by the uniform consistency of $\widehat{W}$ and the orthogonality of $\hat{H}$ and X. This implies that the two terms on the right side of (5.2) are asymptotically independent. Let V denote the limit in probability of $n^{-1}X'WX$. As in Huffer and McKeague (1991), the first term in (5.2) converges in distribution to a Gaussian martingale m with covariation process $\int_0^{\cdot} V^{-1}\,dt$. This is simply the limit of $n^{1/2}(\hat{A}-A)$ in the usual additive risk model in which $\beta = 0$. It follows that $n^{1/2}(\hat{A}-A)$ converges in distribution to $m + \psi(\cdot)\xi$, where m and ξ are independent, ξ is mean zero multivariate normal with variance Σ^{-1}, and $\psi(t) = \int_0^t V^{-1}U\,ds$ where U is the limit in probability of $n^{-1}X'WZ$.

It remains to show that the above argument can be modified to allow for the non-predictability of $\widehat{W}$ when the weights are computed via method II. First consider the last part of (5.1), $n^{-1/2}\int Z'\hat{H}\,dM$, each component of which can be written in the form

$$n^{-1/2}\sum_{i=1}^{n}\int G_i(\tilde{\beta})\,dM_i \tag{5.3}$$

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 (5.3) as

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

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 $n^{1/2}$-consistent and the integrand is predictable and uniformly bounded. The third term is also asymptotically negligible since $\ddot{G}_i(\beta)$ is uniformly bounded for β belonging to a neighbourhood of β_0. Note that we have been using the boundedness of covariates and the assumption that λ is bounded away from zero. A similar argument applies to the first term on the right side of (5.2). We conclude that the asymptotic distribution of $\hat{\beta}$ and $\hat{A}$ is the same for methods I and II.

Appendix: 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

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

Thus for all such b,

$$\begin{aligned} 0 &= E\left[\int \frac{z - b^*(t)'x}{\lambda(t|x,z)}\, dM(t) \int \frac{b(t)'x}{\lambda(t|x,z)}\, dM(t)\right] \\ &= E\left[\delta \frac{(z - b^*(T)'x)}{\lambda(T|x,z)} \frac{b(T)'x}{\lambda(T|x,z)}\right], \end{aligned}$$

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

$$\begin{aligned} b^*(t)' &= E\Big[\frac{zx'}{\lambda^2(t|x,z)}\Big|T, \delta = 1\Big]\Big(E\Big[\frac{xx'}{\lambda^2(t|x,z)}\Big|T, \delta = 1\Big]\Big)^{-1} \\ &= E\Big[\frac{zx'}{\lambda(t|x,z)} 1_{[T \geq t]}\Big]\Big(E\Big[\frac{xx'}{\lambda(t|x,z)} 1_{[T \geq t]}\Big]\Big)^{-1}, \end{aligned}$$

see Sasieni (1992, section 3).

Thus, for a sample of size n,

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

Solving $l^*_\beta = 0$ for β gives (2.3).

We conclude that the estimators $\hat{\beta}$ and $\hat{A}(\cdot)$ discussed in this paper are asymptotically efficient for the semiparametric model (Bickel et. al., 1992).

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics* **2** 1–25. Springer-Verlag, New York.

Aalen O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **8** 907–925.

Barlow, R. E. and Prentice, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75** 65–74.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1992). *Efficient and Adaptive Inference in Semiparametric Models*, Johns Hopkins University Press, Baltimore.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman Hall, London.

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

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman Hall, London.

Cuzick, J., Cooper, E. H., and MacLennan, I. C. M. (1985). The prognostic value of serum β_2 microglobulin compared with other presentation features in myelomatosis. *British J. Cancer* **52** 1–6.

Cuzick, J., De Stavola, B. L., Cooper, E. H., Chapman, C. and MacLennan, I. C. M. (1990). Long-term prognostic value of serum β_2 microglobulin in myelomatosis. *British J. Haematology* **75** 506–510.

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models. *Stoch. Proc. Appl.* **36** 353–370.

Greenwood, P. E. and Wefelmeyer, W. (1991). Efficient estimating equations for nonparametric filtered models. In: *Statistical Inference in Stochastic Processes* **1** (Prabhu and Basawa, eds.) 107–141.

Henderson, R. and Milner, A. (1991). Aalen plots under proportional hazards. *J. Roy. Statist. Soc. Ser. C* **40** 401–410.

Huffer, F. W. and McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Amer. Statist. Assoc.* **86** 38–53.

Lin, D. Y. and Ying, Z. (1992). Semiparametric analysis of the additive risk model. Preprint.

Mau, J. (1986). On a graphical method for the detection of time-dependent effects of covariates in survival data. *Appl. Statist.* **35** 245–255.

Mau, J. (1988). A comparison of counting process models for complicated life histories. *Appl. Stoch. Models Data Anal.* **4** 283–298.

McKeague, I. W. (1988a). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *Contemporary Math.* **80** 139–152.

McKeague, I. W. (1988b). A counting process approach to the regression analysis of grouped survival data. *Stoch. Process. Appl.* **28** 221–239.

Medical Research Council (1984). Analysis and management of renal failure in the fourth MRC myelomatosis trial. *British Medical J.* **288** 1411–1416.

Murphy, S. A. and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model *Stoch. Processes Appl.* **39** 153–180.

Sasieni, P. D. (1992). Information bounds for the additive and multiplicative intensity models. In: *Survival Analysis: State of the Art*, (J. P. Klein and P. K. Goel, eds.) pp. 249–265. Kluwer.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–60.

Zucker, D. M. and Karr, A. F. (1990). Survival data regression analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18** 329–353.

Figure 1

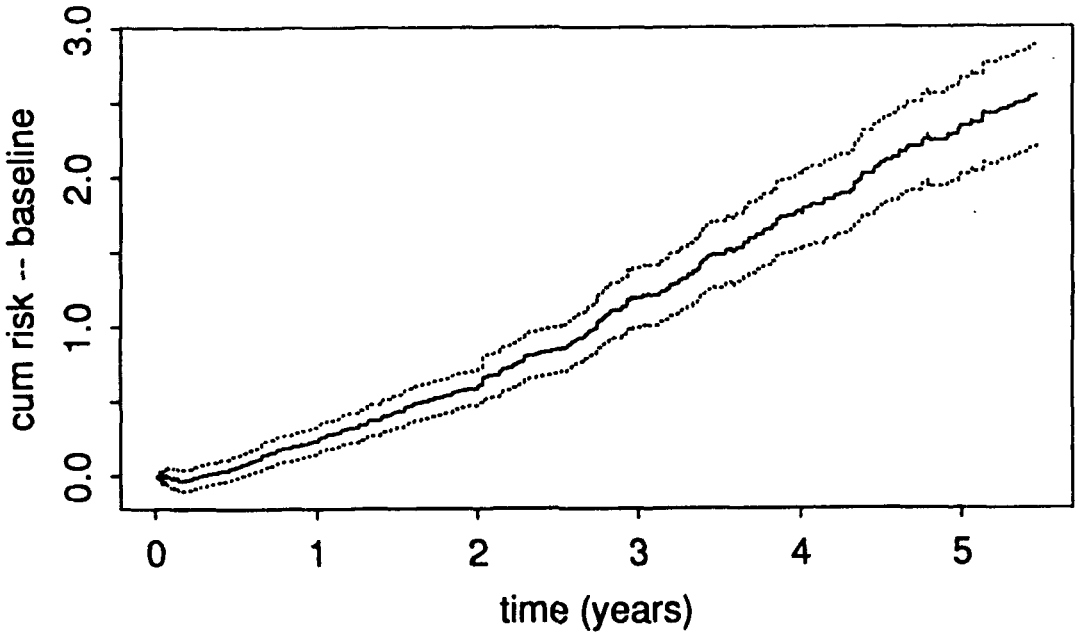

Figure 2

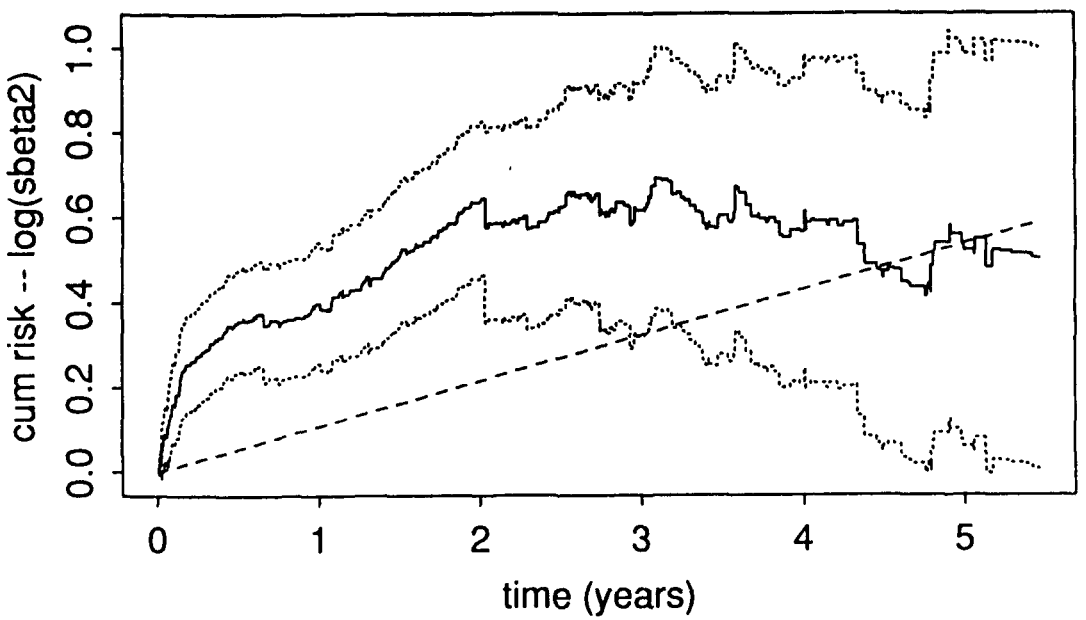

Figure 3

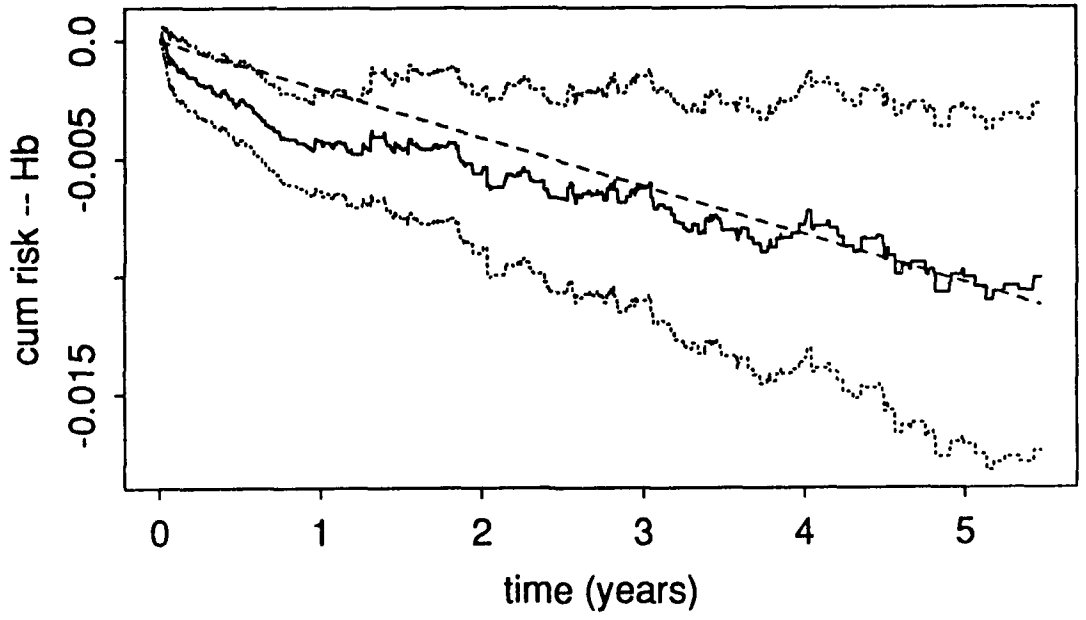

Figure 4

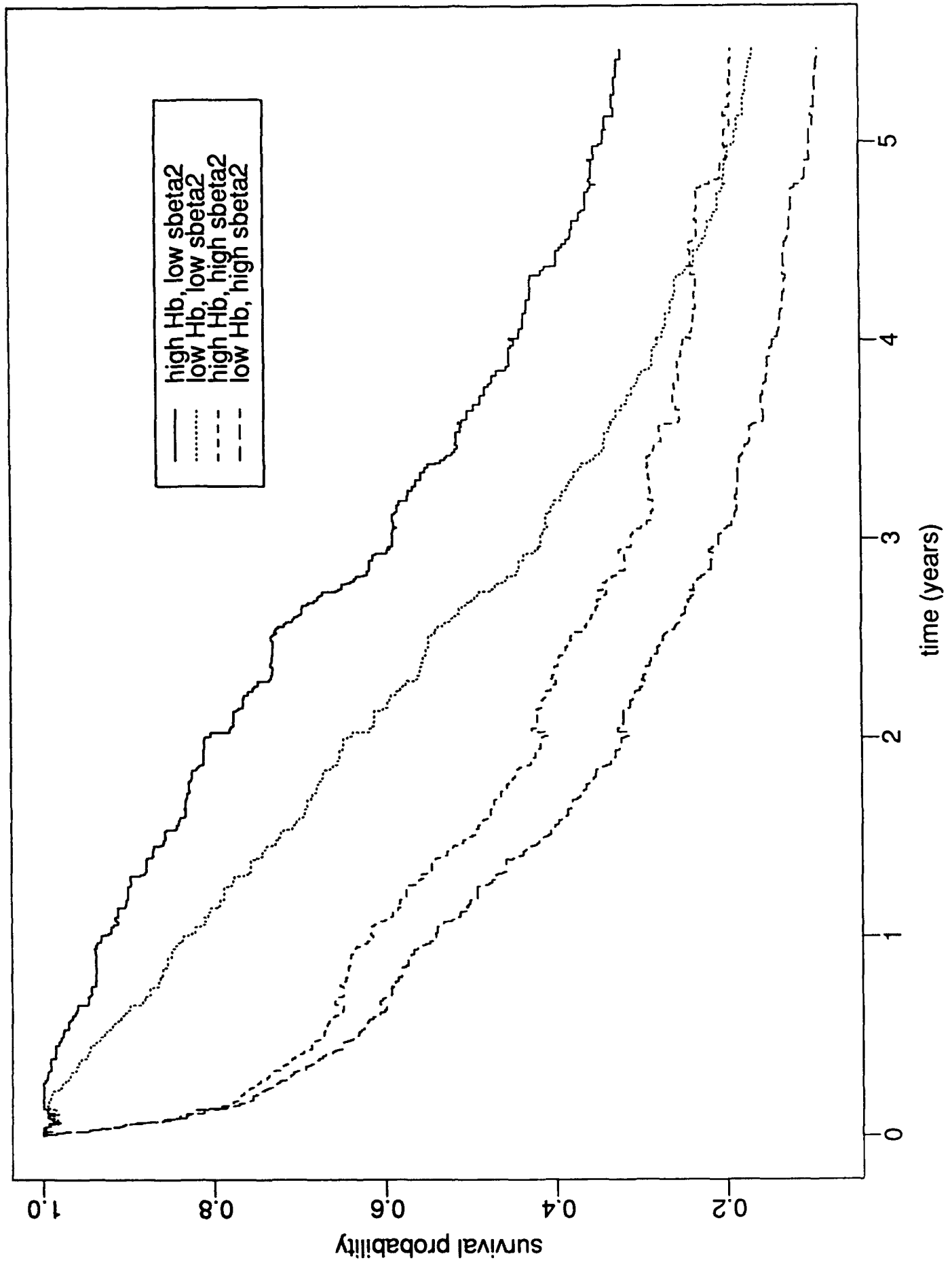
high Hb, low sbeta2
low Hb, low sbeta2
high Hb, high sbeta2
low Hb, high sbeta2
survival probability
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
time (years)

Figure 5

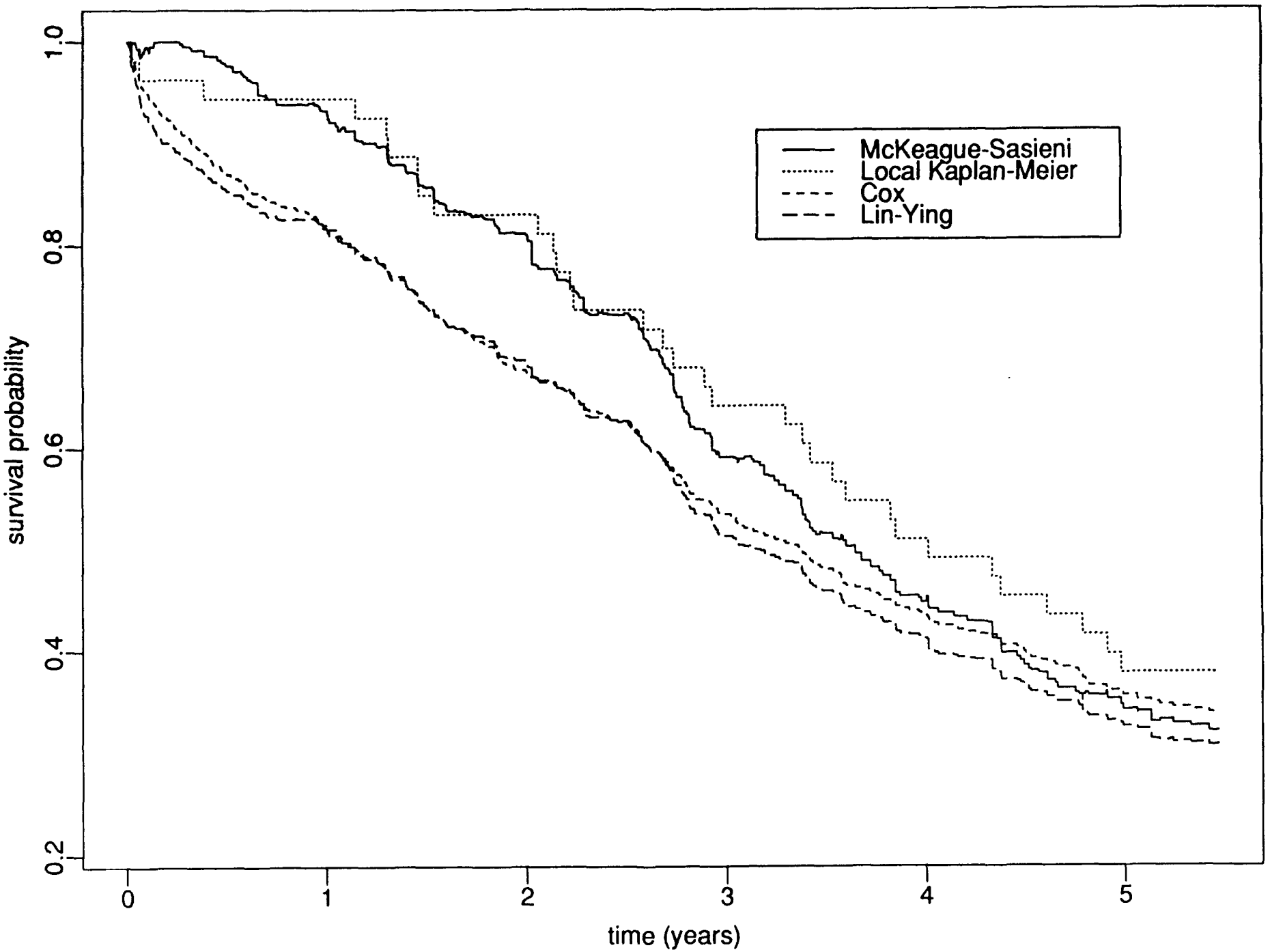
McKeague-Sasieni
Local Kaplan-Meier
Cox
Lin-Ying
survival probability
1.0
0.8
0.6
0.4
0.2
0
1
2
3
4
5
time (years)

Figure 6

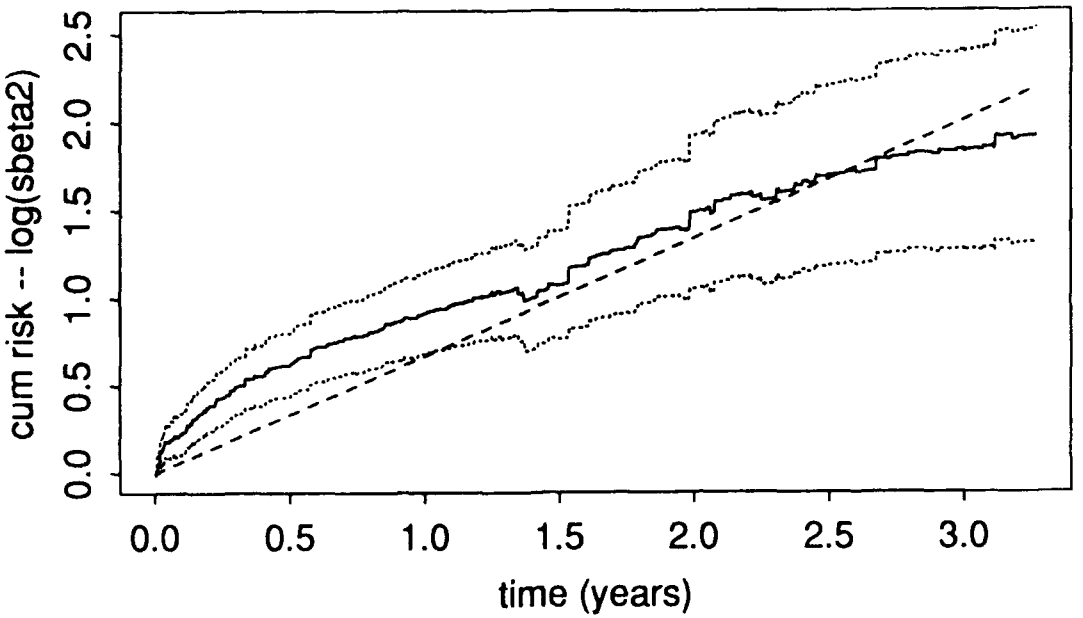

Figure 7