The Annals of Statistics
2001, Vol. 29, No. 5, 1344–1360

A GENERALIZED ADDITIVE REGRESSION MODEL FOR SURVIVAL TIMES [1]

BY THOMAS H. SCHEIKE

University of Copenhagen

We present a non-parametric survival model with two time-scales. The time-scales are equivalent up to a constant that varies over the subjects. Covariate effects are modelled linearly on each time scale by additive Aalen models. Estimators of the cumulative intensities on the two time-scales are suggested by solving approximate local maximum likelihood estimating equations. The local estimating equations necessitate only the choice of one bandwidth. The estimators are provided with large sample properties. The model is applied to data on patients with myocardial infarction, and used to describe the prognostic effect of covariates on the two time scales, time since myocardial infarction and age.

1. Introduction. In many bio-medical applications in survival analysis it is of interest to study the effect of covariates on various time-scales. We consider the situation where multiple time-scales are involved, and focus on the specific situation with two time-scales that are equivalent up to a constant for each individual such as for example follow-up time and age. A class of models where it is relevant to consider multiple time-scales is the the three-state model known as the illness-death model, or the disability model, where the additional time-scale may be duration in the illness state of the model; see Keiding (1991) for a general discussion of these models. Oakes (1995) discussed how multiple time-scales may be combined into a single scale. Previous work has considered semi-parametric survival models where one time-scale is modelled parametrically and the other time-scale is considered non-parametric. An example of this type of analysis with three time-scales applied to diabetes patients can be found in Ramlau-Hansen et al. (1987).

In this paper we present a non-parametric regression approach with two time-scales where each time-scale contribute additively to the mortality. The effect of covariates are modelled by additive Aalen models on each time-scale [Aalen (1980,1989,1993), McKeague (1988), Huffer and McKeague (1991)]. This allows covariates to have effects that vary on two different time-scales. In a motivating example we consider patients that experience myocardial infarction, and aim at predicting the intensity considering the two time-scales age and time since myocardial infarction. As an example, one of the covariates describes the heart function and is allowed to have an effect that varies

Received April 2000; revised May 2001.

[1]Supported by NIH Grant NIH 5 R01 CA 54706-08.

AMS 2000 *subject classifications*. Primary 62N01; secondary 62N02, 62G20

Key words and phrases. Additive Aalen model, counting process, disability model, illness-death model, generalized additive models, multiple time-scales, non-parametric estimation, survival data, varying-coefficient models.

with both age and time since myocardial infarction. The suggested model extends the Generalized Additive Models of Hastie and Tibshirani (1986) to a regression set-up, and we therefore term the models as Generalized Additive Regression Models. Note, however, that the model may be considered as a varying-coefficient model [Hastie and Tibshirani (1993)]. In the survival setting many authors have dealt with generalized additive models. Recent papers are Kooperberg, Truong and Stone (1995), Huang (1999) and Linton, Nielsen and Van de Geer (1999).

By actively utilising that one of the time scales is an ordinary time-scale, unlike markers as a proxy for disease progression, we present a simple non-parametric estimator whose asymptotic properties are derived. One advantage of the proposed methodology is that only one smoothing parameter needs to be chosen.

Section 2 presents the model and some counting process notation. Section 3 gives some local estimating equations that are solved to give simple explicit estimators of the non-parametric effects of the model. Based on these explicit estimators we are able to derive asymptotic results and provide the estimators with asymptotic standard errors. Section 4 considers the special case where some covariates have effects on both time-scales. Two solutions to the identifiability problem are suggested. Section 5 contains an application to data on myocardial infarction. Finally, Section 6 presents some possible extensions.

2. An Aalen model for multiple time-scales. Let $N_i(t),\ i = 1, \dots, n$, be n independent counting processes that do not have common jumps and are adapted to a filtration that satisfy the usual conditions [Andersen, Borgan, Gill and Keiding (1993)]. The counting processes are observed in the time period $[0, \tau]$. We assume that the counting processes have intensities given by

$$\lambda_i(t) = X_i(t)^T \alpha(t) + Z_i(t)^T \beta(t + a_i) \tag{2.1}$$

where $X_i(t) \in \Re^p$, and $Z_i(t) \in \Re^q$ are predictable cadlag covariate vectors, and a_i is a real-valued random variable observed at time $t = 0$. A simple and important submodel is the generalized additive model where $\lambda_i(t) = \alpha(t) + \beta(t + a_i)$. For this model some constraint is needed for one of the non-parametric effects to identify the parameters of the model, we return to this in Section 4 below.

The model contains two time-scales t and a where one of the time-scales is specific to each individual. In the illness-death model, say, t might be time since diagnosis (duration) among subjects that have entered the illness stage of the model and a_i could be the age when the transition to the illness stage occurred, such that $t + a_i$ is the age of the subject.

After introducing some notation we present an estimation procedure that leads to explicit estimators of $A(t) = \int_0^t \alpha(s) ds$ and $\beta(a)$. We derive the asymptotic distribution for these estimators. Based on the cumulative intensity $A(t)$ one may estimate the intensity $\alpha(t)$ by smoothing techniques. We also consider estimation of the cumulative $B(t) = \int_{\tau_a}^t \beta(s) ds$ where τ_a is some lower-limit that depends on the observed range of the second time-scale. The cumulative effects have the advantage compared to $\alpha(s)$ and $\beta(a)$ that they may be used for inferential purposes since a more satisfactory simultaneous convergence can be established for these processes.

2.1. *Notation.* Let $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$ such that $M_i(t) = N_i(t) - \Lambda_i(t)$ are martingales. Let further $N(t) = (N_1(t), \ldots, N_n(t))$ be the n-dimensional counting process, $\Lambda(t) = (\Lambda_1(t), \ldots, \Lambda_n(t))$ be its compensator, such that $M(t) = (M_1(t), \ldots, M_n(t))$ is an n-dimensional martingale, and define matrices $X(t) = (X_1(t), \ldots, X_n(t))^T$ and $Z(t) = (Z_1(t), \ldots, Z_n(t))^T$.

We further need weight matrices and define $W(t) = diag(w_i(t))$ as an $n \times n$ diagonal matrix with weights $w_i(t)$ $i = 1, \ldots, n$.

Define predictable matrices $H_1(s)$ and $H_2(s)$ of dimension respectively, $k \times n$ and $l \times n$. The optional variation process between the martingale integrals $\tilde{M}_1(t) = \int_0^t \tilde{H}_1(s)dM(s)$ and $\tilde{M}_2(t) = \int_0^t \tilde{H}_2(s)dM(s)$ is given as the $k \times l$ matrix

$$[\tilde{M}_1, \tilde{M}_2](t) = \int_0^t \tilde{H}_1(s)\text{diag}(dN(s))\tilde{H}_2(s)^T.$$

Define $[\tilde{M}_1](t) = [\tilde{M}_1, \tilde{M}_1](t)$.

3. Local estimating equations. We now present an estimation procedure that is relatively easy to analyse and provide with asymptotic standard errors since we obtain explicit estimators. The method may further be extended to more general time-scales.

We assume that

$$\lambda_i(t) = X_i(t)^T\alpha(t) + Z_i(t)^T\beta(t + a_i). \tag{3.1}$$

When $(X(s), Z(s))$ has full rank we can identify the effects without imposing constraints on the parameters. If some covariates enter both X and Z we suggest two solutions to solve the identifiability problem in Section 4. We start, however, by assuming that $(X(s), Z(s))$ asymptotically has full rank.

When $t + a_i$ is close to fixed but arbitrary a we have that

$$\lambda_i(t) = X_i(t)^T\alpha(t) + Z_i(t)^T\beta(a) + (t + a_i - a)Z_i(t)^T\beta'(a) + R_i(t) \tag{3.2}$$

where β', β'' are the first and second derivative and $R_i(t) = (1/2)(t + a_i - a)^2 Z_i(t)^T\beta''(a^*_{i,t})$ where $a^*_{i,t}$ lies between $t + a_i$ and a. For fixed a, estimates of $\alpha(\cdot)$ and $\beta(a)$ may be obtained by fitting a kernel weighted version of the above semiparametric model where the remainder term is omitted. McKeague and Sasieni (1994) gave estimators for the semiparametric additive survival model and we use their formulae with kernel weights to account for the amount of information present. A similar idea was used by Li and Doss (1995) to estimate conditional survival based on a fully non-parametric regression model $(\lambda(t, X))$. Based on the local equation (3.2) we can estimate $\beta(a)$ and part of $\alpha(\cdot)$. By fixing a we essentially condition on it and can only estimate $\alpha(\cdot)$ for times s that are observed for the given a. We therefore need some notation to specify what part of $\alpha(\cdot)$ can be estimated based on the local equations as well as notation for the kernel weights. Let the density of a_i be denoted $f(a)$ and define $A(t|a) = \int_0^t I(f(a-s) > 0)\alpha(s)ds$. For fixed a we can estimate $A(t|a)$. Note that $f(a-s)$ is the density for individuals who have age a at time s. Let $W_b(t, a) = diag(w_1(t)K(b, a-(a_1+t)), \ldots, w_n(t)K(b, a-(a_n+t)))$ where $K(b, \cdot)$ is a symmetric kernel function with bandwidth b, and define $\Delta(t, a) = diag((t+a_1-a), \ldots, (t+a_n-a))$, $Z(t, a) = (Z(t), \Delta(t, a)Z(t))$ and combine the parameters into $\nu(a) = (\beta(a)^T, \beta'(a)^T)^T$. In the remainder of the paper the weights are assumed identical to 1 and can thus be ignored. Define further e_1 to be the matrix that gives $\beta(a)$ when multiplied by $\nu(a)$, that is, $\beta(a) = e_1\nu(a)$.

Then by (2.3) and (2.4) in McKeague and Sasieni (1994) estimators of $\nu(a)$ and $A(t|a)$ are given by

$$\hat{\nu}(a) = G(\tau, a)^{-1} \int_0^\tau J(s, a)Z(s, a)^T H(s, a)dN, \tag{3.3}$$

$$\hat{A}(t|a) = \int_0^t J(s, a)X(s, a)^-(dN(s) - Z(s, a)\hat{\nu}(a)ds) \tag{3.4}$$

where $X(s, a)^- = (X(s)^T W_b(s, a)X(s))^{-1}X(s)^T W_b(s, a)$, $H(s, a) = W_b(s, a) - W_b(s, a)X(s)X(s, a)^-$, $J(s, a)$ is one when the inverse in $X(s, a)^-$ exists and 0 otherwise, and $G(t, a) = \int_0^t J(s, a)Z(s, a)^T H(s, a)Z(s, a)ds$.

Using the martingale structure we find that (see the Appendix for additional details)

$$\hat{A}(t|a) - A(t|a) = M_1(t) - C_2(t, a)M_2(\tau) + O_p(1)b^2 + o_p(n^{-1/2}) \tag{3.5}$$

and that

$$\hat{\nu}(a) - \nu(a) = G(\tau, a)^{-1}M_2(\tau) + O_p(1)b^2 + o_p(n^{-1/2}) \tag{3.6}$$

where $C_1(t, a) = \int_0^t X(s, a)^- Z(s, a)ds$, $C_2(t, a) = C_1(t, a)G(\tau, a)^{-1}$ and

$$M_1(t) = \int_0^t X^-(s, a)dM(s),$$

$$M_2(t) = \int_0^t J(s, a)Z(s, a)^T H(s, a)dM(s).$$

The asymptotic description of the estimators $\hat{A}(t|a)$ and $\hat{\beta}(a) = e_1\hat{\nu}(a)$ is given in the following proposition.

PROPOSITION 1. *Assume*:

(i) $K(\cdot)$ *is a compact kernel*, $b \to 0$ *as* $n \to \infty$;

(ii) *Subjects are independent and identically distributed with bounded covariates such that* $(X(s), Z(s))$ *has full rank with probability tending to one*;

(iii) *The intensities are twice continuously differentiable*;

(iv) a_i *has distribution given by a continuously differentiable density* $f(\cdot)$, *and there exist* $s \in [0, \tau]$ *such that* $f(a-s) > 0$.

Then $\sqrt{nb}(\hat{\beta}(a) - \beta(a))$ *converges toward a normal distribution with mean* 0, *and a variance that may be estimated consistently by the* $q \times q$ *matrix*

$$(nb)e_1 G(\tau, a)^{-1}[M_2](\tau)G(\tau, a)^{-1}e_1^T, \tag{3.7}$$

plus a bias term of order $\sqrt{nb}b^2 O(1)$. *Further,* $\sqrt{nb}(\hat{A}(t|a) - A(t|a))$ *is asymptotically equivalent to a Gaussian process with mean* 0, *and a covariance matrix that is estimated consistently by*

$$\begin{aligned}(nb)([M_1](t) &+ C_2(t, a)[M_2](\tau)C_2(t, a)^T \\ &-[M_1, M_2](t)C_2(t, a)^T - C_2(t, a)[M_2, M_1](t)),\end{aligned} \tag{3.8}$$

plus a bias term of order $\sqrt{nb}b^2 O(1)$.

REMARKS. (i) Asymptotically equivalent estimators of the asymptotic variances (3.7) and (3.8) are obtained by replacing $Z(s, a)$ by $Z(s)$ in the formulae.

(ii) The achieved rate is equivalent to the optimal one dimensional rate.

To improve the performance and rate of the cumulative estimator of $\int_0^t \alpha(s)ds$ it seems natural to combine the estimates for different a. We also introduce the process of cumulative effects of $\beta(a)$ on some relevant range of values for a denoted $[\tau_a, \tau_b]$. This process is useful for inferential purposes. Define

$$\hat{A}(t) = \int_0^t \int_{\tau_a}^{\tau_b} E(s, a)\hat{A}(ds|a)da, \tag{3.9}$$

$$\hat{B}(a) = \int_{\tau_a}^{a} \hat{\beta}(c)dc, \tag{3.10}$$

where $E(s, a)$ is predictable and satisfies that $\int_{\tau_a}^{\tau_b} E(s, a)da = 1$ for all s and is used for weighting the increments

$$\begin{aligned}\hat{A}(ds|a) = J(s, a)(X(s)^T W_b(s, a)X(s))^{-1}X(s)^T W_b(s, a)(dN(s) \\ - Z(s, a)\hat{\nu}(a)ds).\end{aligned}$$

In the later example the weights were also used to threshold away increments were only little information is present. An alternative to weighting together the increments is to weight together the cumulatives, $A(t|a)$, such that the estimator of $A(t)$ becomes $\int_{\tau_a}^{\tau_b} E(t, a)\hat{A}(t|a)da$. In the rest of the paper we consider (3.9) because there is some numerical indication this procedure had superior properties, due to its ability to threshold out extreme increments for low information areas.

By (3.6) it follows that

$$\hat{B}(a) - B(a) = \int_0^\tau L(s, a)dM(s) + O_p(1)b^2 + o_p(n^{-1/2})$$

where $L(s,a) = \int_{\tau_a}^{a} e_1 G(\tau, c)^{-1} J(s,c) Z(s,c)^T H(s,c)\, dc$, and by (3.5) it follows that

$$\hat{A}(t) - A(t) = \int_0^t Q_1(s) dM(s) - \int_0^\tau Q_2(s,t) dM(s) + O_p(1) b^2 + o_p(n^{-1/2}) \tag{3.11}$$

where $Q_1(s) = \int_{\tau_a}^{\tau_b} X(s,a)^- E(s,a) da$ and

$$Q_2(s,t) = \int_{\tau_a}^{\tau_b} C_1(t,a) G(\tau,a)^{-1} J(s,a) Z(s,a)^T H(s,a) E(s,a) da.$$

Define further the processes $M_3(t) = \int_0^t Q_1(s) dM(s)$ and $M_4(\tau, t) = \int_0^\tau Q_2(s,t) dM(s)$ that are asymptotically equivalent to martingales (see the Appendix).

The asymptotic properties are described by the following proposition.

PROPOSITION 2. *Under the assumptions of Proposition* 1 *and if* $\sqrt{n} b^2 \to 0$ *it follows that* $\sqrt{n}(\hat{B}(a) - B(a))$ *converges toward a Normal distribution with mean* 0 *and a variance that is estimated consistently by*

$$n \int_0^\tau L(s,a) diag(dN(s)) L(s,a)^T. \tag{3.12}$$

Further if for all $s \in [0, \tau]$ *there exist a such that* $f(a-s) > 0$ *it follows that* $\sqrt{n}(\hat{A}(t) - A(t))$ *is asymptotically equivalent to a Gaussian process on* $[0, \tau]$ *with mean* 0, *and a covariance matrix that is estimated consistently by*

$$n([M_3](t) + [M_4(\cdot,t)](\tau) - [M_3, M_4(\cdot,t)](t) - [M_4(\cdot,t), M_3](t)) \tag{3.13}$$

REMARKS. (i) $\sqrt{n}(\hat{B}(a) - B(a))$ converges toward a Gaussian process under additional regularity conditions, stated in the proof of Proposition 2, that ensures tightness. Therefore, one may use the cumulative process as a basis for inferential purposes.

(ii) Asymptotically equivalent estimators of the asymptotic variance estimators (3.12) and (3.13) are obtained by replacing $Z(s,a)$ by $Z(s)$ in the formulae.

(iii) Tests and simultaneous confidence bands may be based on this Proposition. Note, however, that simulation techniques must be applied to construct simultaneous confidence intervals and tests.

(iv) Based on $\hat{A}(t)$ one can estimate $\alpha(s)$.

4. Generalized additive regression. When some covariates enter both the X and Z design some adaptation is needed. We denote the common covariates as $V(s)$, and let $X_D(s)$ and $Z_D(s)$ be the additional distinct covariates. The model is

$$\lambda_i(t) = X_{Di}(t)^T \alpha_x(t) + Z_{Di}(t)^T \beta_z(t + a_i) + V_i(t)^T \alpha_v(t) + V_i(t)^T \beta_v(t + a_i) \tag{4.1}$$

where $X_{Di}(t) \in \Re^p$, $Z_{Di}(t) \in \Re^q$ and $V_i(t) \in \Re^r$ are predictable cadlag covariate vectors, and a_i is a real-valued random variable with density. We assume that $(X_D(s), Z_D(s), V(s))$ has full rank.

Fitting the local model

$$\begin{aligned}\lambda_i(t) = X_{Di}(t)^T\alpha_x(t) + Z_{Di}(t)^T\beta_z(a) + (t + a_i - a)Z_{Di}(t)^T\beta_z'(a)\\ + V_i(t)^T(\alpha_v(t) + \beta_v(a)) + (t + a_i - a)V_i(t)^T\beta_v'(a) + R_i(t)\end{aligned} \tag{4.2}$$

as above with the full rank design $X(t) = (X_D(t), V(t))$ and

$$Z(t) = (Z_D(t), \Delta(t, a)Z(t), \Delta(t, a)V(t))$$

gives estimates of $(A_x(t|a), A_{vv}(t|a))$ and $\beta_z(a)$ where the subscripts refer to different designs and where $A_{vv}(t|a) = \int_0^t I(f(a - s) > 0)(\alpha_v(s) + \beta_v(a))ds$. We call the estimators $\hat\beta_z(a)$, $\hat A_x(t|a)$ and $\hat A_{vv}(t|a)$, and have that $\hat A(t|a) = (\hat A_x(t|a), \hat A_{vv}(t|a))$. These estimators are thus based on (3.3) and (3.4) and may be cumulated over a by (3.9). Proposition 1 and Proposition 2 are valid for these estimators. Let e_v be a matrix such that $e_v\hat A(t|a) = \hat A_{vv}(t|a)$.

The identifiability problem now needs to be solved to separate the effects $\alpha_v(t)$ and $\beta_v(a)$. We present two approaches to do this. One simple approach assumes that the pair (t, a) is observed in a rectangle. This assumption will be unrealistic in many applications. Another approach gives a more complicated asymptotic analysis but only assumes that when a_1 is close to a_2 then we observe many of the same time-points.

4.1. *Rectangular support region.* The simplest situation is when we restrict attention to a rectangular support region where $f(a - s) > 0$ for all $a \in [\tau_a, \tau_b]$ and $s \in [0, \tau]$. Then assuming $\int_0^\tau \alpha_v(s)ds = 0$ gives estimators

$$\hat\beta_v(a) = \hat A_{vv}(\tau|a)\frac{1}{\tau},$$

$$\hat A_v(t|a) = \hat A_{vv}(t|a) - \hat A_{vv}(\tau|a)\frac{t}{\tau},$$

$$\hat A_v(t) = \int_0^t \int_{\tau_a}^{\tau_b} E(s, a)\hat A_v(ds|a)da.$$

These estimators can be written as martingale integrals by the use of (3.11) and (3.5) and

$$\begin{aligned}\hat A_v(t) - A_v(t) = \int_0^t \int_{\tau_a}^{\tau_b} E(s, a)(\hat A_{vv}(ds|a) - A_{vv}(ds|a))da\\ - \int_0^\tau \int_{\tau_a}^{\tau_b} E_c(t, a)(\hat A_{vv}(ds|a) - A_{vv}(ds|a))da\end{aligned} \tag{4.3}$$

where $E_c(t, a) = (1/\tau)\int_0^t E(s, a)ds$. The first term on the right hand side of (4.3) can be written as

$$\int_0^t \tilde Q_1(s)dM(s) - \int_0^\tau \tilde Q_2(s, t)dM(s) + O(1)b^2 + o_p(n^{-1/2})$$

where $\tilde{Q}_1(s) = e_v Q_1(s)$ and $\tilde{Q}_2(s) = e_v Q_2(s)$ (see (3.11)).

The second term on the right hand side of (4.3) has a similar martingale expression with $\tilde{Q}_1^c(s,t) = \int_{\tau_a}^{\tau_b} e_v X(s,a)^- E_c(t,a)da$ and

$$\tilde{Q}_2^c(s,t) = \int_{\tau_a}^{\tau_b} e_v C_1(\tau,a) G(\tau,a)^{-1} J(s,a) Z(s,a)^T H(s,a) E_c(t,a) da$$

such that

$$\begin{aligned}\hat{A}_v(t) - A_v(t) &= \int_0^t \tilde{Q}_1(s) dM(s) - \int_0^\tau \tilde{Q}_1^c(s,t) dM(s) \\ &\quad - \int_0^\tau (\tilde{Q}_2(s,t) - \tilde{Q}_2^c(s,t)) dM(s) \\ &\quad + O(1)b^2 + o_p(n^{-1/2}).\end{aligned}$$

The optional variation based on the martingale integrals can be used to estimate the variance. Note, however, that the integrands are not all predictable. $\hat{A}(\tau)$ is not equal to 0 because we weighted the increments, rather than the cumulatives $(A(t|a))$. The estimator $\int E(t,a)\hat{A}(t|a)da$ is simpler to analyse and satisfies that it is equal to 0 in τ and results in a similar estimator when numerical problems for low information areas do not occur [the inverse of $X^T(s) W_b(s,a) X(s)$ may become large].

Finally, with

$$\hat{B}_v(a) = \int_{\tau_a}^{a} \hat{\beta}_v(c) dc$$

it follows that

$$\begin{aligned}(\hat{B}_v(a) - B_v(a)) &= \frac{1}{\tau} \int_0^\tau \int_{\tau_a}^{a} (\hat{A}_{vv}(dt|a) - A_{vv}(dt|a)) da \\ &= \int_0^\tau P_1(s,a) dM(s) - \int_0^\tau P_2(s,a) dM(s) + O(1)b^2 + o_p(n^{-1/2})\end{aligned}$$

where $P_1(s,a) = \frac{1}{\tau} \int_{\tau_a}^{a} X(s,u)^- du$ and

$$P_2(s,a) = \frac{1}{\tau} \int_{\tau_a}^{a} C_1(\tau,u) G(\tau,u)^{-1} J(s,u) Z(s,u)^T H(s,u) du.$$

Under assumptions similar to those made in Proposition 2 it may be derived that the cumulative processes are asymptotically Gaussian with a variance that is estimated by using the martingale structure similarly to what was done in the previous section.

4.2. *Smoothness of support region.* Often the observations will not make the simple solution presented above available when solving the identifiability problem because $f(a-s)$ is not positive for all combinations of a and s in their domains. An alternative solution is based on the observation that

$$d\beta_v(a_1,a_2) = \frac{\int_0^\tau J(s,a_1)J(s,a_2)h(s,a_1,a_2)(\hat{A}_{vv}(ds|a_1)-\hat{A}_{vv}(ds|a_2))}{\int_0^\tau J(s,a_1)J(s,a_2)h(s,a_1,a_2)ds}$$

$$\xrightarrow{\text{p}} \beta_v(a_1)-\beta_v(a_2)$$

if $J(s,a_1)J(s,a_2)$ converges uniformly to 1 on some part of the time interval and $h(s,a_1,a_2)$ is some weight function such that $\int_0^\tau h(s,a_1,a_2)ds = 1$ for all a_1 and a_2. Note, that the difference is only estimable where $J(s,a_1)J(s,a_2)$ is different from 0. Therefore, generally, it makes sense to have a_1 and a_2 close to each other.

We assume that $\beta_v(\tau_a)$ is zero to solve the identifiability problem. Now, given a grid of points $\tilde{a}_0=\tau_a,\dots,\tilde{a}_n=\tau_b$ we have that

$$\hat{\beta}_v(\tilde{a}_i) = \sum_{1\le j\le i} d\hat{\beta}_v(\tilde{a}_i,\tilde{a}_{i-1})$$

The estimator of $\beta_v(a_i)$ has a martingale expansion that may be derived as in the previous section. The asymptotic variance may be estimated using this expression.

5. Application to the TRACE study. The TRACE study group [see, e.g., Jensen et al. (1997)] has collected information on 6000 consecutive patients with myocardial infarction (AMI) with the aim of studying the prognostic importance of various risk factors on mortality. At the time of entry (time of AMI) the patients had various risk factors recorded such as age, gender (male=1), congestive heart failure (CHF) (present=1) and ventricular fibrillation (VF) (present=1). Some risk factors were expected to have strongly time-varying effects, in particular ventricular fibrillation. Two time-scales are relevant for studying the mortality. We chose time since AMI as the primary time-scale (t) and age as the secondary time-scale (a). We then avoid the difficult choice of a smoothing parameter on the primary time-scale where it is known that survival changes quite dramatically. Age effects are expected to vary more smoothly and therefore it is much easier to chose a reasonable bandwidth for the age time-scale. The total number of deaths in a 3 year period after entering the study was 2020, and of these, 649 took place within the first month. Figure 1 shows the death times for the patients on the two time scales. A rectangular support region can not be ruled out for these data. We estimate the age effects for ages between $[60, 90]$, and consider the primary time-scale in the time-period from $[0, 3]$.

We considered the following intensity model:

$$\lambda_i(t) = \alpha_1(t)+\alpha_2(t)\text{SEX}_{i1}+\alpha_3(t)\text{VF}_{i1}+\alpha_4(t)\text{CHF}_{i1}$$
$$+\beta_1(t+a_i)+\beta_2(t+a_i)\text{SEX}_{i1}+\beta_3(t+a_i)\text{VF}_{i1}+\beta_4(t+a_i)\text{CHF}_{i1}$$

where a_i is the age of the ith subject at the time of entry. The primary interest centers on estimating the effects of VF and CHF while it was necessary to

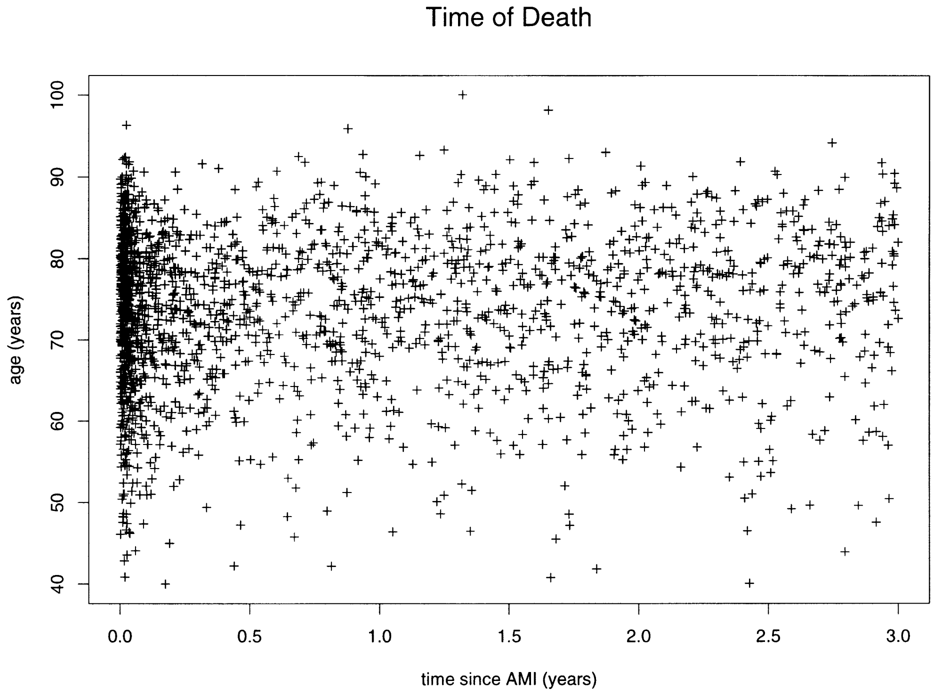

FIG. 1. *Time of death given as a function of time since AMI and age.*

correct for different mortality due to gender and age. The model allows that effects vary with time since AMI and age. To identify the model we assume that $\int_0^3 \alpha_j(s)ds = 0$ for all j.

Estimates of $A(t)$ and $\beta(a)$ were obtained using the formulae presented in the previous sections. A grid of ages were chosen as 20 equidistantly spaced points between 60 and 90 $(\tilde{a}_1, \ldots, \tilde{a}_m)$. We used the following predictable weight function

$$E(s, \tilde{a}_j) = \frac{I(e_i(b, a_j, s) > 10)e_i(b, a_j, s)}{\sum_j \sum_i I(e_i(b, a_j, s) > 10)e_i(b, a_j, s)}$$

where $e_i(b, a_j, s) = \sum_i K(b, \tilde{a}_j - (a_i + s))$ and $I(e_i(b, a_j, s) > 10)$ is the indicator of the event $e_i(b, a_j, s) > 10$. The bandwidth was 5 years and a Tukey kernel was used in the weight function. The indicator was introduced to ignore ill-behaved estimates based on only a few observations, and improved the performance of the estimator. Recall that the weight function is used for weighing together the increments of $A(t|a)$.

Figure 2 shows the age effects with 95 % pointwise variability intervals that were identified based on an assumption of rectangular support. The variability intervals are based on (3.7) and thus ignore the bias term. Note that the age effect is essentially the same for men and women and that both CHF and VF appears to have effects that vary with age. A formal test may be based on the

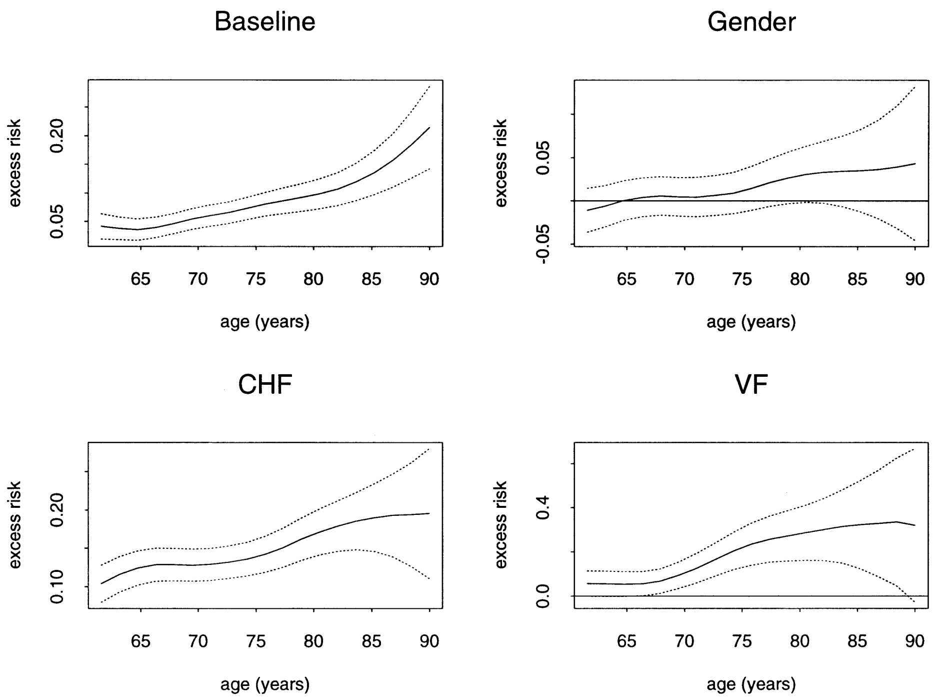

FIG. 2. *Age-varying effects of covariates with* 95% *pointwise variability intervals.*

joint asymptotic distribution of the cumulative effects but has not been carried out. Figure 3 shows the cumulative time-effects with 95 % pointwise confidence intervals. Note the strongly time-varying effect of VF on the time-scale time since AMI. The presence of VF results in a highly increased mortality the first couple of months and then the effect wears off. Note, that all cumulative estimators start and end at 0 due to the chosen solution of the identifiability problem. The effects of both VF and CHF are significantly time-varying and level off to constant non-significant levels after the first couple of months. A formal test may be based on the asymptotic description.

6. Extensions. Analysis of the semi-parametric submodel where some of the covariate effects do not vary with time can be carried out along the same lines. Note, that the second time-axis in principle could be any marker that is thought relevant for the subject matter, whereas it is used actively in the local estimating equations and in the martingale derivations of the proofs that we have one ordinary time-scale. In principle the local estimating equations may be extended to additional time-scales.

When some covariate has effects on both time-scales we suggested a simple procedure to separate the effects. Even when effects had to be separated we were able to obtain asymptotic results but further research is needed into this area.

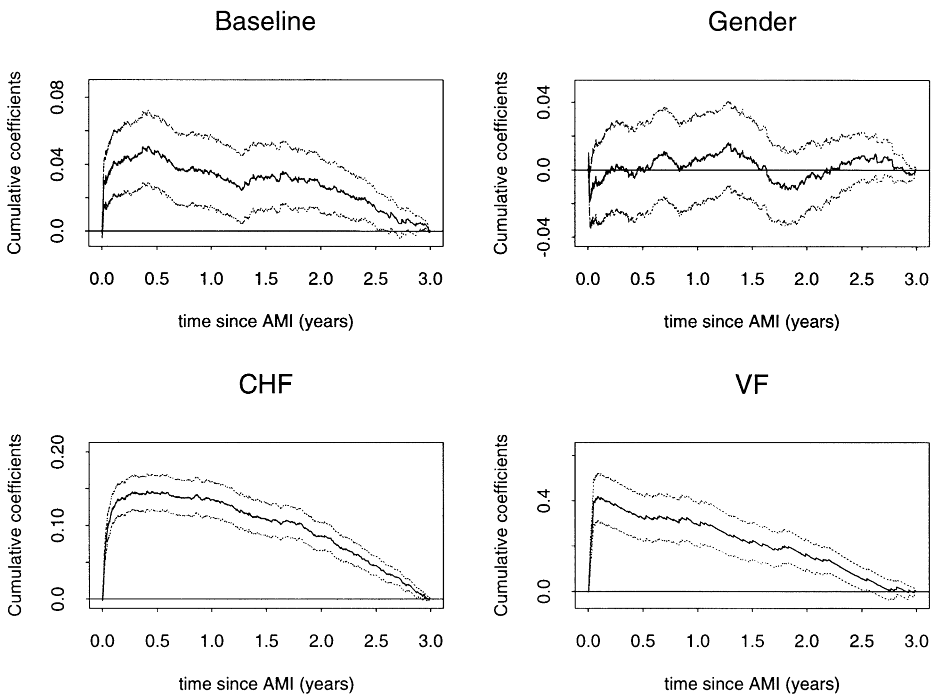

FIG. 3. *Time-varying cumulative effects with* 95% *pointwise confidence intervals.*

Following the arguments of Linton, Nielsen and Van de Geer (1999) the asymptotic description of the estimates based on the local estimating equations may be used as the starting point for a back-fitting algorithm based on the usual Aalen estimating equations. It is possible that such an estimator is efficient, but further research is needed to study this issue.

APPENDIX

SKETCH OF PROOFS. The semi-parametric additive intensity model was studied by McKeague and Sasieni (1994), and the proofs for Proposition 1 resemble theirs. The proofs of McKeague and Sasieni (1994) use various techniques from McKeague (1988) and Huffer and McKeague (1991).

Recall

$$M_1(t,a) = \int_0^t X^-(s,a)dM(s),$$

$$M_2(t,a) = \int_0^t J(s,a)Z(s,a)^T H(s,a)dM(s),$$

$$X^-(s,a) = (X(s)^T W_b(s,a)X(s))^{-1}X(s)^T W_b(s,a).$$

PROOF OF PROPOSITION 1. For simplicity we take W as the identity matrix. A martingale expansion yields that

$$\begin{aligned}\hat{\nu}(a)-\nu(a) = {} & G(\tau,a)^{-1}M_2(t)\\ & +G(\tau,a)^{-1}\int_0^\tau J(s,a)Z(s,a)^T H(s,a)\Delta(t,a)\Delta(t,a)Z_L(s)\tilde{\beta}''(s)ds\\ & +o_p(n^{-1/2})\end{aligned}$$

$Z_L(t)$ is an $n\times nq$ matrix and $\tilde{\beta}''(s)=(\beta''(a^*_{1,s})^T,\dots,\beta''(a^*_{n,s})^T)^T$ such that the i'th element of $Z_L(t)\tilde{\beta}(t)$ is $Z_i(t)^T\beta''(a^*_{i,s})$ and $a^*_{i,s}$ lies between $s+a_i$ and a (note that this argument in reality is q-dimensional). The bias term is seen to be $O(1)b^2$ by using the fact that the second derivative is bounded on a compact set. The martingale term can be shown to converge in distribution if the predictable variation converges in probability, a Lindeberg condition is satisfied, and $G(s,a)^{-1}$ converges in probability. To see that $G(s,a)^{-1}$ converges in probability we note that with $\mu_2(K)=\int u^2K(u)du$,

$$\begin{aligned}(1/n)Z^TW_bZ &= f(a-s)E(Z_i^T(s)Z_i(s)|a_i=a-s)+o_p(1),\\ (1/n)(\Delta Z)^TW_bZ &= b^2\mu_2(K)D_s(f(\cdot)E(Z_i^T(s)Z_i(s)|a_i=\cdot))(a-s)+o_p(b^2),\\ (1/n)(\Delta Z)^TW_b(\Delta Z) &= b^2\mu_2(K)f(a-s)E(Z_i^T(s)Z_i(s)|a_i=a-s)+o_p(b^2),\end{aligned}$$

uniformly in s by the law of large numbers combined with smoothness. $D_s(g)$ is the derivative of g with respect to s. Similar expressions are obtained when X replaces Z in the above formulae. Based on these an asymptotic expression for the inverse of $G(s,a)$ may be computed by using that for non-singular A,

$$A=\begin{pmatrix}A_{11}+o_p(1) & b^2A_{12}+o_p(b^2)\\ b^2A_{21}+o_p(b^2) & b^2A_{22}+o_p(b^2)\end{pmatrix}$$

the inverse can be written as

$$\begin{aligned}A^{-1} &= \begin{pmatrix}\tilde{A}_{11} & \tilde{A}_{12}\\ \tilde{A}_{21} & \tilde{A}_{22}\end{pmatrix}\\ &= \begin{pmatrix}A_{11}^{-1}+b^2A_{11}^{-1}A_{12}D^{-1}A_{21}A_{11}^{-1}+o_p(1) & -A_{11}^{-1}A_{12}D^{-1}+o_p(1)\\ -D^{-1}A_{21}A_{11}^{-1}+o_p(1) & b^{-2}D^{-1}+o_p(b^{-2})\end{pmatrix}\end{aligned}$$

where $D=A_{22}-b^2A_{21}A_{11}^{-1}A_{12}$. It is seen that $\tilde{A}_{11}=A_{11}^{-1}+o_p(1)$, $\tilde{A}_{12}=A_{11}^{-1}A_{12}A_{22}^{-1}+o_p(1)$, and $\tilde{A}_{22}=b^{-2}A_{22}^{-1}+o_p(b^{-2})$.

It therefore follows that $\hat{\beta}(a)-\beta(a)$ is asymptotically equivalent to

$$\left(\int_0^\tau J(s,a)Z(s)^TH(s,a)Z(s)\,ds\right)^{-1}\int_0^\tau J(s,a)Z(s)^TH(s,a)dM(s)$$

The Lindeberg condition follows from the assumption of bounded covariates, and the predictable variation multiplied by nb equals

$$(nb)\int_0^\tau J(s,a)Z(s)^TH(s,a)\text{diag}(\lambda_i(t))(Z(s)^TH(s,a))^T\xrightarrow{p}\Sigma$$

where the convergence follows as above.

Similarly,

$$\hat{A}(t|a) - A(t|a) = \int_0^t (1 - J(s,a))\alpha(s)ds + M_1(t)$$
$$- \int_0^t X(s,a)^- Z(s,a)(\hat{\nu}(a) - \nu(a))ds$$
$$- \int_0^t X(s,a)^- \Delta(t,a)\Delta(t,a) Z_L(s)\tilde{\beta}''(s)ds$$
$$= M_1(t) - \int_0^t X(s,a)^- Z(s,a)(\hat{\nu}(a)$$
$$-\nu(a))ds + O_p(1)b^2 + o_p(n^{-1/2})$$
$$= M_1(t) - C_1(t,a)G(\tau,a)^{-1} \int_0^\tau J(s,a)Z(s,a)^T H(s,a)dM(s)$$
$$+O_p(1)b^2 + o_p(n^{-1/2})$$

where $C_1(t,a) = \int_0^t X(s,a)^- Z(s,a)ds$. It follows that $C_1(t,a)G(\tau,a)^{-1}$ converges in probability toward a limit $c(t,a)$. The bias term is bounded by $O(1)b^2$. The martingale term is seen to converge as above, and combined with the martingale expansion for $(\hat{\nu}(a) - \nu(a))$, Proposition 1 follows. □

PROOF OF PROPOSITION 2.

$$\int_{\tau_a}^d (\hat{\beta}(a) - \beta(a))da = \int_0^\tau \int_{\tau_a}^d e_1 G(\tau,a)^{-1} J(s,a)Z(s,a)^T$$
$$\times H(s,a)\, da\, dM(s) + O(1)b^2$$
$$= \int_0^\tau L(s,d)dM(s) + O(1)b^2$$

where $L(s,d) = \int_{\tau_a}^d e_1 G(\tau,a)^{-1} J(s,a)Z(s,a)^T H(s,a)\, da$. Again, an asymptotically equivalent expression is given by replacing $Z(s,a)$ by $Z(s)$ in the formula and omitting e_1. For simplicity we consider this asymptotically equivalent expression in the remainder of the proof. Let $g(\tau,a) = lim_p(ne_1 G(\tau,a)^{-1})$ and define the predictable $\tilde{L}(s,d) = \int_{\tau_a}^d g(\tau,a)J(s,a)Z(s)^T H(s,a)\, da$. The finite dimensional convergence follows from Martingale convergence, if the non-predictable L process can be replaced by a predictable version. Considering the one-dimensional case. We want that $n^{-1/2} \sup |\int_0^\tau (nL(s,d) - \tilde{L}(s,d))dM(s)| \overset{p}{\to} 0$. The difference can be written as $n^{-1/2} \sum_i \int \delta_i dM_i$ where subscript i refers to the ith element of the vectors and $\delta_i = (L_i - \tilde{L}_i)$. Now, δ_i is a sum of two terms $d_{i,1} = \int (e_1 nG(\tau,a)^{-1} - g(\tau,a))J(s,a)K(b, a-(a_i + s))da Z_i(s)^T$ and $d_{i,2} = \int (e_1 nG(\tau,a)^{-1} - g(\tau,a))J(s,a)K(b, a - (a_i+s))Z(s)^T W_b(s,a)X(s)(X(s)^T W_b(s,a)X(s))^{-1} da X_i(s)^T$. Considering the first term it can be written as

$$\begin{aligned}
&\left|\sum_i n^{-1/2}\int_0^\tau \int (ne_1 G(\tau,a)^{-1} - g(a))K(b,a-(a_i+s))da\, Z_i(s)\, dM_i(s)\right| \\
&\quad \le \sum_i n^{-1/2}\int \left|(ne_1 G(\tau,a)^{-1} - g(a))\int_0^\tau K(b,a-(a_i+s))Z_i(s)dM_i(s)\right| da \\
&\quad \le \sup_a \left|(ne_1 G(\tau,a)^{-1} - g(a))\right|(\tau_b-\tau_a) \\
&\qquad \times \sup_a \left| n^{-1/2}\sum_i \int_0^\tau K(b,a-(a_i+s))Z_i(s)dM_i(s)\right|.
\end{aligned}$$

The last martingale term is seen to be bounded in probability when multiplied by b. The term therefore converges to 0 because of the rate of the first term. Similarly it is seen that the second term converges to 0 in probability. Therefore it follows that L can be replaced by a predictable approximation, $\tilde{L}$, that can be studied by the martingale convergence theorem. A different approach for solving this predictability problem was described in Nielsen (1999).

To prove tightness one may show the moment condition $nE(\int_{d_1}^{d_2}(\hat{\beta}(a) - \beta(a))da)^2 \le C|d_1-d_2|^2$, and consider only the one dimensional case. To avoid difficulties with the inverse of $(X^T W_b X)$ being to large when computing the mean we work with a bounded version of this inverse as in McKeague (1988), and show as in McKeague that this substitution of the inverse gives the same asymptotics. Using the martingale expression we find that

$$\begin{aligned}
&nE\left(\int_{d_1}^{d_2}(\hat{\beta}(a)-\beta(a))da\right)^2 \\
&\quad = \frac{1}{n}E\int_0^\tau (\tilde{L}(t,d_1)-\tilde{L}(t,d_2))\,\mathrm{diag}(\lambda_i(t)\,dt)(\tilde{L}(t,d_1)-\tilde{L}(t,d_2))^T
\end{aligned}$$

and since covariates and intensities are bounded it suffices to show that

$$\begin{aligned}
\frac{1}{n}\sum_i(\tilde{L}_i(t,d_1)-\tilde{L}_i(t,d_2))^2 &\le C_1|d_1-d_2|^2\, sup_a|g(\tau,a)| \\
&\le C|d_1-d_2|^2
\end{aligned}$$

where $\tilde{L}_i(t,d_1)$ is the ith element of $\tilde{L}(t,d_1)$ and where the last inequality follows if $g(\tau,a)$ is continuous.

Now, turning to $A(t)$

$$\begin{aligned}
\hat{A}(t) - \int_0^t \alpha(s)ds = &\int_0^t\int_{\tau_a}^{\tau_b} X(s,a)^- E(s,a)\,da\,dM(s) \\
&- \int_0^t\int_{\tau_a}^{\tau_b} X(s,a)^- Z(s,a)(\hat{\nu}(a)-\nu(a))E(s,a)\,da\,ds \\
&+ O(1)b^2 + o_p(n^{-1/2})
\end{aligned}$$

The last term is equal to

$$\begin{aligned}\int_0^t \int_{\tau_a}^{\tau_b} X(s,a)^- Z(s,a) G(\tau,a)^{-1} \int_0^\tau J(u,a) Z(u,a)^T \\ H(u,a)\, dM(u)\, E(s,a) da\, ds \\ = \int_0^\tau \biggl(\int_0^t \int_{\tau_a}^{\tau_b} X(s,a)^- Z(s,a) G(\tau,a)^{-1} J(u,a) Z(u,a)^T \\ \times H(u,a)\, E(s,a)\, da\, ds \biggr) dM(u)\end{aligned}$$

Its asymptotic properties are described by the Martingale convergence theorem if the non-predictable integrands can be replaced by predictable approximations. So is suffices to show a Lindeberg condition and that the predictable variation converges in distribution when multiplied by n. That the nonpredictable process $Q_2(s, t)$ can be replaced by a predictable version follows as before be replacing $C_1(\tau, a)G(\tau, a)^{-1}$ by its limit in probability $c_1(\tau, a)g(\tau, a)^{-1}$. □

Acknowledgments. I am grateful to the TRACE study group for letting me use their data. I would also like to thank two referees and an Associate Editor for their helpful comments.

REFERENCES

AALEN, O. O. (1980). A model for non-parametric regression analysis of counting processes. *Lecture Notes in Statist.* **2** 1–25. Springer, New York.

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

AALEN, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine* **12** 1569–1588.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models. *J. Amer. Statist. Assoc.* **81** 297–318.

HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796.

HUANG, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563.

HUFFER, F. W. and MCKEAGUE, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Amer. Statist. Assoc.* **86** 114–129.

KEIDING, N. (1991). Age-specific incidence and prevalence: A statistical perspective (with discussion). *J. Roy. Statist. Soc. Ser. A* **154** 371–396.

KOOPERBERG, C., STONE, C. and TRUONG, Y. (1991). The L2 rate of convergence for hazard regression. *Scand. J. Statist.* **22** 143–158.

LI, G. and DOSS, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.* **23** 787–823.

LINTON, O., NIELSEN, J. P. and VAN DE GEER, S. (1999). Estimating multiplicative and additive marker dependent hazard functions by backfitting with the assistance of marginal integration. Technical Report 7, Dept. Statistics, Univ. Copenhagen.

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

MCKEAGUE, I. W. and SASIENI, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81** 501–514.

NIELSEN, J. P. (1999). Super efficient hazard estimation. *Biometrika* **86** 227–232.

OAKES, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Anal.* **1** 7–18.

RAMLAU-HANSEN, H., JESPERSEN, N. C. B., ANDERSEN, P., BORCH-JOHNSEN, K. and DECKERT, T. (1987). Life insurance for insulin-dependent diabetics. *Scand. Actuar. J.* 19–36.

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF COPENHAGEN
BLEGDAMSVEJ 3
COPENHAGEN, DK-2200
DENMARK
E-MAIL: ts@biostat.ku.dk