Dept. of Math. University of Oslo
Statistical Research Report No. 9
ISSN 0806-3842 August 2004

Path analysis for survival data with recurrent events

Johan Fosen

Institute of Basic Medical Sciences, University of Oslo / The Norwegian Cancer Society

Ørnulf Borgan

Department of Mathematics, University of Oslo

Harald Weedon-Fekjær

Institute of Basic Medical Sciences, University of Oslo / The Norwegian Cancer Registry

Odd O. Aalen

Institute of Basic Medical Sciences, University of Oslo

Abstract

We propose a method for path analysis of survival data with recurrent events. By applying an additive model for the intensity, concepts like direct, indirect and total effects may be defined in an analogous way as for traditional path analysis. The focus is on understanding how to analyze the effect of a dynamic covariate, e.g. the number of previous events, and at the same ensuring that the effect of a fixed covariate is unbiasedly estimated. Theoretical considerations as well as simulations are presented. A dataset on recurrent tumors in rats is used for illustration.

Keywords: causal analysis, path analysis, dynamic covariates, event history analysis, graphical models, internal covariates, orthogonalization, treatment effect, Aalen's additive regression model

1 Introduction

In spite of the apparent success of the field, survival and event history analysis has had shortcomings which have just recently started to receive a solution. The analysis of time-dependent internal covariates has been a persistent problem. It has been well known that a treatment effect may be wrongly estimated when time-dependent covariates are included, e.g. in a Cox model. The recent revised version of the classical book by Kalbfleisch and Prentice (2002) contains discussion on this point, but no solution is presented. Important contributions have been made in the marginal structural models of Robins and coworkers, see e.g. Robins, Hernán and Brumback (2000). However, our perspective here is different, or more general, since we want to understand the effect of the time-dependent covariate, *per se*, instead of considering it as a nuisance to be corrected for.

Our approach is built upon the classical path analysis which was developed as an extension of linear regression models (Wright, 1934). In the simplest case one analyses a series of regression models corresponding to a causal understanding of the relationships between the variables under study. This approach has found a wide application in many fields and is the basis on which the more recent and far more extensive graphical models have been derived (e.g. Pearl, 2000). The statistical analysis of simple recursive path models is based on linear regression calculations, that is, successive least squares estimation.

In this paper we combine these procedures with the additive, or linear, model for hazard regression that has been developed in the counting process context (Aalen, 1980, 1989). In this model a local least square estimation is performed whenever an event occurs, and the local estimates are added up in an informative cumulative plot. Including a standard path model for the covariates, yields a new path model for covariates and events combined. The algebraic calculations from classical path analysis all transfer over to this new situation because they only depend on the linearity and the least square estimation. The assumption of normally distributed errors is of course not relevant for the hazard part of the model, but is substituted by the machinery of counting processes.

A major advantage of such a path analysis is the distinction between direct, indirect and total effects. Marginal structural models, on the other hand, do

not give such detailed analysis. For the Cox model the analogues of direct, indirect and total effects do not seem to have been developed, no doubt due to the difficulties incurred by nonlinearity. In fact, even for logistic regression models it has been difficult to suggest suitable analogues of direct, indirect and total effects, although there may exist a recent suggestion (Eshima et al, 2001). The difficulty of defining indirect effects for nonlinear models is also discussed by Pearl (2001), who makes some general suggestions, without however relating this to specific examples like the Cox model.

We shall consider the situation where a number of processes with recurrent events are studied. These could be repeated occurrences of disease, repeated awakenings during a night, etc (see e.g. Aalen et al, 2004, for examples). A common approach for analysing such recurrent events has been to use frailty models (Hougaard, 2000). Recently a different approach has been studied, namely to let the number of previous events for an individual be a covariate giving information on the individual specific risk. This was applied in Kalbfleisch and Prentice (2002), Aalen et al (2004), Peña and Hollander (2004), Gandy and Jensen (2004) and Miloslavsky, Keleş and van der Laan (2004). We want to point out possible pitfalls in analyses with covariates like 'the number of previous events'. Nevertheless, we believe this is a useful and important procedure, but one has to be careful to avoid bias. The authoritative textbook of Kalbfleisch and Prentice (2002) uses the number of previous events as covariate (Chapter 9) without sufficient clarification, in our opinion.

What we want to do is to make a joint analysis getting a correct picture both of the effect of fixed covariates, including possibly treatment, and of the dynamic properties of the underlying process with repeated occurrences for each individual. Note that also other dynamic (i.e, dependent on the past) covariates than the number of previous events may be involved. A model with dynamic covariates will be termed a dynamic model, see Aalen et al (2004) for a detailed discussion.

In Section 2 we give a brief presentation of the nonparametric additive hazard model. The extension by Scheike (2002), where the model is presented also as a rate function model, is important here. Usually in counting process theory one studies models for the intensity processes, such that residuals are

martingales. This requires a complete modelling of covariates and past events influencing the present intensity. Because we need to also analyse marginal models, we shall have to consider situations where a more limited amount of information is included in the model. This is precisely Scheike's rate function model, where residuals are not martingales, and where in particular special methods for variance estimation and testing must be applied.

Section 3 gives some theoretical results on the relationship between various additive models. A study of a data set on recurrent tumours in rats, which is used to illustrate the procedures, is presented in Section 4. Here we also give a simulation study. The approach through local path analysis is explained in Section 5. A simulation study and discussion on variance estimation is included in Section 6. Our approach could also handle more complex problems, and we give some comments on this towards the end.

2 The additive hazard model

Consider a situation where n individuals are followed over a time period, and the times of occurrence of a recurrent event are registered. Let $\tilde{N}_i(t)$ denote the counting process representing the number of events for individual i by follow-up time t. Since an individual may experience several events, the process $\tilde{N}_i(t)$ will take integer values $0, 1, 2, 3, ...$ For individual i we have the covariates $Z_{i1}(t), ..., Z_{ip}(t)$. These may be fixed or depend on time, and the time-dependent covariates at time t are allowed to depend on $\tilde{N}_i(s)$ for $s < t$. A statistical model for this situation is obtained by specifiying how the hazard or intensity process $\alpha_i(t)$ of $\tilde{N}_i(t)$ depends on the covariates; cf. Kalbfleisch & Prentice (2002, section 9.1). We will assume that the intensity is given by the additive hazard model:

$$\alpha_i(t) = \beta_0(t) + \beta_1(t)Z_{i1}(t) + ... + \beta_p(t)Z_{ip}(t) \tag{1}$$

(cf. Aalen 1980, 1989). Here the regression functions $\beta_j(t)$ are arbitrary in t and describe the possibly time-varying effects of the covariates. Estimation in the additive model usually focuses on the cumulative regression functions $B_j(t) = \int_0^t \beta_j(s)ds$.

2.1 Data and estimation

The individuals may be followed over different periods of time. We let $Y_i(t) = 1$ if individual i is under observation "just before" time t, and let $Y_i(t) = 0$ otherwise. We also introduce the process $N_i(t) = \int_0^t Y_i(s)d\tilde{N}_i(s)$ counting the observed number of the event for individual i, and assume that censoring is independent in the sense of Kalbfleisch & Prentice (2002), see also Andersen et al. (1993). Then the intensity process of the observed counting process $N_i(t)$ takes the form

$$\lambda_i(t) = Y_i(t)\alpha_i(t) \tag{2}$$

Note that our set-up allows for both right-censoring and left-truncation. However, covariates depending on previous events in the underlying counting process $\tilde{N}_i(t)$ can cause problems for left-truncated data.

By combining (1) and (2) we obtain

$$\lambda_i(t) = \mathbf{W}_i(t)\boldsymbol{\beta}(t),$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), ..., \beta_p(t))'$ is the vector of regression functions, and $\mathbf{W}_i(t) = Y_i(t)(1, Z_{i1}(t), ..., Z_{ip}(t))$. By standard results for counting processes, we then have

$$dN_i(t) = \lambda_i(t)dt + dM_i(t) = \mathbf{W}_i(t)\boldsymbol{\beta}(t)dt + dM_i(t), \tag{3}$$

where the $M_i(t)$ are martingales.

In order to write the last expression in vector form, we let $\mathbf{N}(t) = (N_1(t), ..., N_n(t))'$ be the vector of counting processes, the corresponding vector of martingales is $\mathbf{M}(t) = (M_1(t), ..., M_n(t))'$, and $\mathbf{B}(t) = (B_0(t), B_1(t), ..., B_p(t))'$ is the vector of cumulative regression functions, meaning that $dB_j(t) = \beta_j(t)dt$. Then we may write

$$d\mathbf{N}(t) = \mathbf{W}(t)d\mathbf{B}(t) + d\mathbf{M}(t), \tag{4}$$

where $\mathbf{W}(t)$ is the matrix with $\mathbf{W}_i(t)$ as the i'th row. Note that this has the form of a linear model with $d\mathbf{N}(t)$ as response, $\mathbf{W}(t)d\mathbf{B}(t)$ as the systematic

component, and $d\mathbf{M}(t)$ as the noise term. Ordinary least-squares regression then gives

$$d\hat{\mathbf{B}}(t) = \left[\mathbf{W}(t)'\mathbf{W}(t)\right]^{-1}\mathbf{W}(t)'d\mathbf{N}(t) \tag{5}$$

when $\mathbf{W}(t)$ has full rank. We introduce $J(t)$ as the indicator of $\mathbf{W}(t)$ having full rank, and the least-squares generalized inverse

$$\mathbf{W}^{-}(t) = \left[\mathbf{W}(t)'\mathbf{W}(t)\right]^{-1}\mathbf{W}(t)'.$$

Summing up the increments of (5) over the event times where estimation is meaningful, we obtain the estimator

$$\hat{\mathbf{B}}(t) = \int_0^t J(s)\mathbf{W}^{-}(s)d\mathbf{N}(s) = \sum_{T_k \le t} J(T_k)\mathbf{W}^{-}(T_k)\Delta\mathbf{N}(T_k)$$

where $T_1 < T_2 < \cdots$ are the event times and $\Delta\mathbf{N}(T_k)$ is the increment of $\mathbf{N}(T_k)$, being a vector with one for the component corresponding to the individual experiencing an event at T_k and zero otherwise.

We will usually center each of the covariates at every time T_k where we are estimating, that is subtracting the mean of the covariate over the individuals at risk at T_k. The reason is that $\beta_0(t)$ in (1) then can be interpreted as the intensity of an "average" individual, whereas $\beta_0(t)$ otherwise would be the intensity of an individual having value zero for every covariate.

2.2 Variance estimation

If we have a correctly specified model, the vector $d\mathbf{M}(t)$ in (4) is a martingale increment. Introducing $\mathbf{B}^*(t) = \int_0^t J(u)d\mathbf{B}(u)$ we can write

$$\hat{\mathbf{B}}(t) - \mathbf{B}^*(t) = \int_0^t J(u)\mathbf{W}^{-}(u)d\mathbf{M}(u),$$

showing that $\hat{\mathbf{B}}(t) - \mathbf{B}^*(t)$ is martingale. An (essentially) unbiased estimator of the covariance matrix of $\hat{\mathbf{B}}(t)$ is given by the optional variation process of the martingale $\hat{\mathbf{B}}(t) - \mathbf{B}^*(t)$, i.e. by

$$\begin{aligned}\widehat{\mathrm{VAR}}\ \hat{\mathbf{B}}(t) &= \left[\hat{\mathbf{B}} - \mathbf{B}^*\right](t) = \int_0^t J(u)\mathbf{W}^{-}(u)\mathrm{diag}(d\mathbf{N}(u))\mathbf{W}^{-}(u)' \\ &= \sum_{T_k \le t} \mathbf{W}(T_k)^{-}\mathrm{diag}\left(\Delta\mathbf{N}(T_k)\right)\mathbf{W}^{-}(T_k)',\end{aligned}$$

cf. Andersen et al. (1993).

For a misspecified model, the process $\hat{\mathbf{B}}(t)-\mathbf{B}^*(t)$ will fail to be a martingale, and the estimator of the covariance matrix given above may be biased. Scheike (2002) proposed a robust variance estimator based on modelling the rate instead of intensity. The rate differs from the intensity by the following assumptions: whereas the intensity is defined conditionally on the complete history, the rate is defined only given the identity of the individuals at risk and their covariates at time t (or rather just prior to t). Thus the modelling of rates is based on fewer assumptions than the modelling of intensities. Scheike's estimator of the covariance matrix is given by

$$\widetilde{\mathrm{VAR}}\ \hat{\mathbf{B}}(t) = \sum_{i=1}^{n} \hat{\mathbf{Q}}_i(t)\hat{\mathbf{Q}}_i(t)'$$

where

$$\hat{\mathbf{Q}}_i(t) = \int_0^t \left(\mathbf{W}(s)'\mathbf{W}(s)\right)^{-1} \mathbf{W}_i(s)' \left(dN_i(s) - \mathbf{W}_i(s)d\hat{\mathbf{B}}(s)\right).$$

As defined earlier $\mathbf{W}_i(t)$ is the row of $\mathbf{W}(s)$ corresponding to individual no. i. We note that $\hat{\mathbf{Q}}_i(t)$ is the cumulative weighted residual process up to time t for individual i, the weight reflecting the size of the individual's covariates. The variance estimator is the sum of squares of these individual cumulative weighted residuals.

2.3 Ridge regression

When using ordinary least-squares regression, we will sometimes encounter a singular matrix in (5) when either very few individuals are at risk or when, in the beginning a dynamic covariate has identical values for every individual at risk. To avoid the singularity problem we will use ridge regression at times T_k where we encounter a singular matrix.

The idea of ridge regression is to solve the problem of sparse data by putting constraints on the parameter values, thus shrinking the estimates. We minimise $\sum_i \{dN_i(t) - \mathbf{W}_i(t)d\mathbf{B}(t)\}^2 + \eta(t)\sum_j dB_j(t)^2$, where $\eta(t)$ is the ridge coefficient, instead of the usual sum of squared residuals. Note that in our context $\eta(t)$ may be a predictable stochastic process, since the need for performing ridge regression

may change over time dependent on what happens in the counting processes. The ridge estimator for the increment of the regression function becomes

$$d\hat{\mathbf{B}}(t) = \left[\mathbf{W}(t)'\mathbf{W}(t) + \eta(t) \cdot \mathbf{I}\right]^{-1} \mathbf{W}(t)' d\mathbf{N}(t),$$

where $\mathbf{I}$ is the identity matrix of appropriate dimension (here $p+1$). We note that the estimator differs from the usual least-squares estimator only by adding, for any time t, a constant to the diagonal of $\mathbf{W}(t)'\mathbf{W}(t)$.

The parameter $\eta(t)$ can be different for each of the $p+1$ parameters, that is substituting $\eta(t)\cdot\mathbf{I}$ with a diagonal matrix with diagonal elements $\eta_0(t), \eta_1(t), ..., \eta_p(t)$. In practice this is seldom done due to the already fairly large challenges in choosing an appropriate size of the *scalar* $\eta(t)$. However, in our situation where singularity of $\mathbf{W}(T_k)'\mathbf{W}(T_k)$ often happens because of identical values of a dynamic covariate at the start of the study, it would be sufficient to let $\eta_i(t)$ be zero except for the ridge coefficients corresponding to dynamic covariates.

As with other shrinkage methods, ridge regression reduces the variance, providing us with a more stable estimate. There is, however, a trade-off between stability and bias: the variance decreases and the bias increases, both monotonically, with the value of the ridge coefficient. In ordinary regression setting with no censoring, it can be shown (Gruber 1998, theorem 3.2.1) that the mean squared error as a function of $\eta(t)$ will never have its minimum for $\eta(t) = 0$, hence ridge regression is always potentially superior to ordinary least squares regression. However, although there exists clever methods for choosing $\eta(t)$, the choice will often at least partly rely on a subjective assessment of the appropriate degree of shrinkage. If the regression coefficients are very dependent on the exact choice of $\eta(t)$, this indicates an instability in the data, and the results should be interpreted with caution.

3 Connection between various additive models

We shall present here some theoretical considerations, looking at the consistency between various additive models. This is especially important in the path analysis where additive models are viewed at several different levels. The first two cases show that additivity (or linearity in the covariate) is preserved un-

der marginalization. Notice that such a result is not valid for most nonlinear models, like the Cox model. Case 3 concerns the relationship between a frailty model and a dynamic model. This turns out to be more complex and linearity is not preserved.

Throughout this section we consider the situation without censoring so that there is no difference between $\lambda(t)$ and $\alpha(t)$, see (2).

Case 1. Assume the covariates Z_1 and Z_2 are independent random variables and that we have the following true model for the intensity process:

$$\lambda_T(t) = \beta_0(t) + \beta_1(t)Z_1 + \beta_2(t)Z_2$$

Assume now that Z_2 is unknown. The rate function (not intensity process any more!) with Z_1 as the only covariate may be found by the following informal argument:

$$\begin{aligned}\frac{1}{dt}\mathrm{E}(N(t+dt) - N(t) \mid Z_1) &= \frac{1}{dt}\mathrm{E}(\mathrm{E}(N(t+dt) - N(t) \mid Z_1, Z_2)| \, Z_1) \\ &= \mathrm{E}(\beta_0(t) + \beta_1(t)Z_1 + \beta_2(t)Z_2 \mid Z_1) \\ &= \beta_0(t) + \beta_1(t)Z_1 + \beta_2(t)\mathrm{E}(Z_2)\end{aligned}$$

Hence, the marginal model is still linear in Z_1 with the same regression function $\beta_1(t)$, and we can write the rate function as:

$$\lambda_M(t) = \beta_0^*(t) + \beta_1(t)Z_1$$

Notice that we would also get an additive model if Z_1 and Z_2 are correlated and normally distributed (due to the linear conditional mean), but then the regression function of Z_1 would be changed.

Case 2. We shall now start with a dynamic model and see that linearity is preserved under marginalization. Let $N(t-)$ be the number of events in the process prior to t, and assume the following model:

$$\lambda_D(t) = \beta_0(t) + \beta_1(t)Z + \beta_2(t)N(t-)$$

Define the function $f_Z(t) = E(N(t)|Z)$. Following the argument above, the marginal rate function of $N(t)$ given only Z is given as:

$$\begin{aligned}\frac{1}{dt}\mathrm{E}(N(t+dt) - N(t) \mid Z) &= \mathrm{E}(\beta_0(t) + \beta_1(t)Z + \beta_2(t)N(t-) \mid Z) \\ &= \beta_0(t) + \beta_1(t)Z + \beta_2(t)\mathrm{E}(N(t) \mid Z)\end{aligned}$$

This yields the differential equation

$$f'_Z(t) = \beta_0(t) + \beta_1(t)Z + \beta_2(t)f_Z(t),$$

with the initial condition $f_Z(0) = 0$ corresponding to $N(0) = 0$. The solution of this linear differential equation is straightforward and we skip the details. Noting that the marginal rate function equals $f'_Z(t)$, the following expression is found for it:

$$\begin{aligned}\lambda_{DM}(t) &= \beta_0(t) + \beta_2(t)\int_0^t \beta_0(v)\exp(\int_v^t \beta_2(u)du)\,dv \\ &\quad + Z\left\{\beta_1(t) + \beta_2(t)\int_0^t \beta_1(v)\exp(\int_v^t \beta_2(u)du)\,dv\right\}.\end{aligned}$$

Hence, the rate function is still linear in Z although the coefficient has changed. Note that if, for example, $\beta_1(t)$ and $\beta_2(t)$ are both positive, then the coefficient of Z is larger in the marginal than in the dynamic model, confirming results demonstrated in the example and the simulation in the next section.

As an example, let $\beta_0(t) = 0$, $\beta_1(t) = 1$ and $\beta_2(t) = 1$. Then

$$\lambda_{DM}(t) = Z(\exp(t) - 1).$$

Hence the regression function of Z equals 1 in the full model case, and $\exp(t)-1$ in the marginal case, which clearly demonstrates the large difference in the estimated influence of Z which there may be between a full and a marginal model.

Case 3. We start with the model

$$\lambda_T(t) = \beta_0(t) + \beta_1(t)Z_1 + \beta_2(t)Z_2 \tag{6}$$

with independent Z_1 and Z_2, but now we want to see what is the effect when instead we use Z_1 and $N(t-)$ as covariates, which is an alternative when Z_2 is unobserved. This is a typical use of a dynamic model, as a substitute for a model with unknown frailty components. Assume for simplicity that $\beta_0(t) = 0$ and that $\beta_1(t) = \beta_1$ and $\beta_2(t) = \beta_2$ are constants. Given Z_1 and Z_2 the process $N(t)$ is a homogenous Poisson process with rate $\lambda_T(t)$. Assuming $N(t-) = k$

we have by a somewhat informal argument:

$$\begin{aligned}\frac{1}{dt}P(dN(t) &= 1 \mid Z_1,\ \text{past}) = \frac{1}{dt}\frac{P(dN(t)=1,\ \text{past} \mid Z_1)}{P(\text{past} \mid Z_1)} \\ &= \frac{1}{dt}\frac{E(P(dN(t)=1,\ \text{past} \mid Z_1, Z_2) \mid Z_1)}{EP(\text{past} \mid Z_1, Z_2) \mid Z_1)} \\ &= \frac{E((\beta_1 Z_1 + \beta_2 Z_2)^{k+1}\exp(-t(\beta_1 Z_1 + \beta_2 Z_2)) \mid Z_1)}{E((\beta_1 Z_1 + \beta_2 Z_2)^{k}\exp(-t(\beta_1 Z_1 + \beta_2 Z_2)) \mid Z_1)}\end{aligned}$$

Now, assume that the Laplace transform of Z_2 is $L(s)$, that is:

$$L(s) = E(\exp(-sZ_2))$$

Introduce:

$$\varphi(s,x) = L(s)\exp(-s\,x) = E(\exp(-s(x+Z_2)))$$

and note that

$$\phi(s,x,n) = \frac{\partial^n}{\partial}\varphi(s,x) = (-1)^n E((x+Z_2)^n \exp(-s(x+Z_2)))$$

Inserted above this gives the following rate in a dynamic model:

$$\begin{aligned}P(dN(t) &= 1 \mid Z_1,\ \text{past}) = \frac{-\beta_2^{k+1}\,\phi(t\,\beta_2, \beta_1 Z_1/\beta_2, k+1)}{\beta_2^k\,\phi(t\beta_2, \beta_1 Z_1/\beta_2, k)} \\ &= \beta_2 \frac{-\phi(t\beta_2, \beta_1 Z_1/\beta_2, k+1)}{\phi(t\,\beta_2, \beta_1 Z_1/\beta_2, k)} \qquad (7)\end{aligned}$$

This is clearly not a linear model in Z_1, so additivity is not preserved in this case. However, the function might still be approximately linear in many cases. As an example, let Z_2 have an exponential distribution with expectation 1, and assume $\beta_1 = \beta_2 = 1$, $t = 1$ and $k = 2$. Then the function in (7) with Z_1 as argument is plotted in Figure 1 and shown to be approximately linear. In simulations below we shall show that an additive dynamic model fits well to data generated from a model of the type in (6).

4 lllustrations of dynamic covariates

The issues studied shall be demonstrated through a dataset and through simulations. In particular we shall show the underestimation of a treatment effect that may occur when dynamic covariates are included.

4.1 Tumour data

Gail et al. (1980) have described a data set concerning the development of mammary cancer. For 76 female rats, injection of a carcinogen (an agent producing cancer) for mammary cancer were followed by retinyl acetate (cancer prevention) for an initial period of 60 days. The subcohort of 48 rats still being cancer free after the initial period were randomly divided into a treatment group of 23 animals receiving continued cancer prevention, and a control group of 25 animals. All rats were then examined for tumours twice a week until day 122 after randomization.

Let $N(t)$ denote the total number of mammary tumours for a rat by time t where the date of randomization is defined as time zero (note that on the figures the time start at the point of injection). We analyse the data with the additive hazard model using the following covariates:

- Covariate 1: whether in treatment group or in control group (1=treatment, 2=control)
- Covariate 2: the number of previous tumours in the rat until time $t-$, divided by elapsed time since randomization.
- Covariate 3: whether more than 20 days has passed since the last occurrence of a tumour in the rat (1=yes, 0=no).

To avoid singularity problems due to covariate 3, we start the estimation at time 20 after randomization. In order to analyze these data we can use a marginal model, containing only covariate 1. Presumably this gives a correct estimation of the treatment effect, since the treatment was decided by randomization. If at the same time, however, we want to understand the structure of the underlying processes, we should also include covariates 2 and 3 which are dynamic (and then also internal) as described in section 1. From the discussion of Case 2 in Section 3, we can expect an underestimation of the treatment effect in a dynamic model. This is confirmed by our analysis. The upper right panel of Figure 2 shows that the cumulative regression function of the treatment covariate in the marginal model (with only treatment as covariate) reaches approximately the value 2.8 at the final time 122 (182 after injection). On the

other hand, the treatment effect in the full dynamic model (all covariates included) is only about 1.6 at the same time (middle right panel). This means that after the inclusion of the dynamic covariates in the model, we underestimate the treatment effect by approximately 40%.

The phenomenon of underestimation also occurs in a Cox analysis. Table 1 shows the result for the marginal model and the full dynamic model. The underestimation of the treatment effect, as measured by the regression coefficient, is 29%.

Standardized residual processes, e.g. Aalen et al. (2004), are shown in the left panels of Figure 3. The upper panel is for the marginal model and the lower for the dynamic model. The residuals should remain within -2 and +2 if the model used for analysing data is adequate for the true pattern since then they are approximately standard Gaussian distributed at any fixed time point. Both plots have some residuals exceeding +2, but this tendency is smaller when we use a dynamic model, especially when some time has passed. This is in accordance with the fact that the dynamic covariates will not catch the data pattern properly until some events have occurred. The right panels show the mean and standard deviation of the standardized residual processes, and we see that the standard deviation is increasing with time to above 2 in the marginal model, clearly revealing that there are patterns in the data which the marginal model fails to catch. However, in the dynamic model the standard deviation is almost constant in time and equal to 1, showing that the dynamic model fits well.

From the analysis of these data, we see that we have a marginal model that does not fit the data well, but apparently gives a correct picture of the treatment effect, and a dynamic model that gives a much better description of the data as a whole. The treatment effect is very different in the two models, and the question remains how to combine the two analyses to get a true picture of the effect of all covariates, and how the apparent inconsistency between the two models can be resolved.

The issues discussed here shall be further demonstrated in a simulation, before we in Section 5 present the path analysis which resolves the issue.

4.2 Simulation model

Whereas the analysis of the tumour data in the previous section showed that underestimation of treatment effect can be a problem in a model with dynamic covariates, we will in this section through a simulation model try to study when and to what extent this happens.

When describing our simulation, we will use the terms 'simulation model' and 'analysing model'. These terms must not be confused. The former represents by definition "the truth" since the data are generated under this model. Once the data are generated, we are going to analyse them as being real data, thus assuming no knowledge of the true model. We analyse the data by the analysing model, which is our guess of the true model.

We shall use a simple simulation model, mimicking the situation where there is an unknown random effect. More precisely, we shall assume an intensity process $\lambda_T(t) = \beta_0(t) + \beta_1(t)Z_1 + \beta_2(t)Z_2$ where Z_1 corresponds to the known and Z_2 to the unknown covariate. For the statistician the data generation process is unknown, with the knowledge of the variable Z_1 only, and with a qualified guess that there is some additional unmeasured heterogeneity. In the analysing model the statistician uses an additive model with the dynamic covariate $N(t-)$, that is the previous number of events, as an observable substitute for the unknown element. As seen in Section 3, Case 3, the two additive models considered are not consistent and the estimate of the intensity process will be expected to be biased. However, in the simulation we shall carry out residual analysis which shows that the dynamic model still fits the data well, so the inconsistency is only slight and of little practical importance.

More precisely, our simulation model is

$$\lambda_T(t) = 0.5 + kZ_1 + (1-k)Z_2, \tag{8}$$

where k is between 0 and 1. For the Z's we choose three different distributions, uniform between 0 and 1, exponential with parameter 1, and gamma with shape parameter 2 and scale parameter 1. For all situations considered, we generate n=40 processes from the simulation model (8). Two analysing models are considered, the dynamic model

$$\lambda_D(t) = \beta_0^D(t) + \beta_1^D(t)Z_1 + \beta_2^D(t)N(t-) \tag{9}$$

and the marginal model

$$\lambda_M(t) = \beta_0^M(t) + \beta_1^M(t)Z_1. \tag{10}$$

One question is whether $\beta_1^D(t)$ differs much from the "true" effect of Z_1 measured by k, and we shall see that this depends on the value of k.

First we shall consider whether the analysing model with the dynamic covariate $N(t-)$ gives a reasonable fit to the data. Carrying out the estimation we make martingale residual plots as described in Aalen et al (2004). Figure 4 shows the mean and standard deviation of the standardized residual processes for the marginal analyzing model (10) as well as the dynamic model (9) when the covariates are gamma distributed. For the marginal model the standard deviation of the standardized residuals is clearly above 1 for values of k in the lower range, that is, when the unknown covariate has considerable influence. Hence, as expected, the marginal model does not yield a good fit in this case. For the dynamic model the standard deviation is close to 1 for all values of k, indicating a good fit. Figure 5 shows the individual standardized martingale residual processes for the true model with $k = 0.5$ for the two analyzing models. The residuals supports our claim that the dynamic model gives a good fit when the true model is (8).

Next we consider whether applying the dynamic analysing model gives a correct estimate of regression parameter of Z_1. In fact, a considerable underestimation is found dependent on the value of k. Table 2 shows the amount of underestimation for different distributions of the covariates.

As demonstrated in Section 3 a correct estimate of the effect of Z_1 can be found from the marginal model containing only Z_1. However, we are interested in a joint analysis containing both Z_1 and the additional random or dynamic effects. These issues will be sorted out in a path analysis in the next section. Note also that the marginal model destroys the martingale property of the residuals as demonstrated above. Scheike (2002) has provided a robust variance estimator for this case.

5 Local path analysis

We now introduce our version of path analysis. For a basic introduction to standard path analysis, see (Bollen 1998). We perform path analysis at each event time, getting a dynamic picture of the direct, indirect and total effects.

In Section 2 we used the symbol $B(t)$ for a cumulative regression function and $dB(t)$ for its increment. In this section we will compare regression functions from different nested models, and two nested models will in general give us different regression functions. In order to compare regression functions, we will also use $\Theta(t)$ to denote a cumulative regression function. Some models give us identical regression functions, in which case we will use the same symbol for the regression functions in both models.

Throughout we will consider the situation where we have one fixed covariate Z_1 corresponding to treatment, and in addition possibly the dynamic covariate $N(t-)$. We shall assume that all covariates are centered, that is the mean values for the individuals at risk are subtracted. Note that this centering will change over time with a changing risk set. Let $\mathbf{Z}_1^{c,t}$ denote a vector consisting of centered values of Z_1 for individuals at risk at time t, and of zeroes for individuals not at risk. Correspondingly, define $\mathbf{N}^c(t-)$ as the vector of centered values of $N_i(t-)$ or of zeroes for those not at risk.

5.1 Estimation

We first consider the marginal model with only the fixed covariate. A path diagram of this model is given in Figure 6. In the usual model form we can write the model as

$$dN_i(t) = \left\{ d\Theta_0(t) + d\Theta_1(t) Z_{i1}^{c,t} \right\} Y_i(t) + dM_i^*(t). \tag{11}$$

Since the marginal model specifies the rate function in the sense of Scheike (2002) instead of the intensity process, the residual $dM_i^*(t)$ is not necessarily a martingale increment, cf. Section 2.2. The estimator $d\hat{\Theta}_1(t)$, derived as in Section 2.1, will be a correct estimator of the treatment effect.

If we also include the dynamic covariate $N_i^c(t-)$, we get a "naive" dynamic model whose path diagram is given in Figure 7. The model can be written as

$$dN_i(t) = \left\{ dB_0(t) + dB_1(t) Z_{i1}^{c,t} + dB_2(t) N_i^c(t-) \right\} Y_i(t) + dM_i(t). \tag{12}$$

We shall here assume that the description of the intensity process through the covariates $Z_{i1}^{c,t}$ and $N_i^c(t-)$ is a complete one, so that $M_i(t)$ is a martingale. The estimate $d\hat{B}_1(t)$ will usually be smaller than $d\hat{\Theta}(t)$, as we saw in Section 3 (case 2) and in Section 4.2. The intuitive reason for the underestimation is that individuals with a large value of $Z_{i1}^{c,t}$ will tend to have a larger $N_i^c(t-)$, and then some of the difference that actually is due to $Z_{i1}^{c,t}$, will in (12) be accounted for by $N_i^c(t-)$.

5.2 Orthogonal covariates

Two vectors $\mathbf{x}_1 = (x_{11}, ..., x_{n1})'$ and $\mathbf{x}_2 = (x_{12}, ..., x_{n2})'$ are orthogonal if the inner product is zero, that is $\sum_{i=1}^n x_{i1}x_{i2} = 0$. A geometrical interpretation is that $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal vectors in a n-dimensional space.

In ordinary least square regression, it can be shown that if we consider two nested models differing by one covariate, and where this covariate is orthogonal to the other covariates in the model, then the regression coefficients of the common covariates will be estimated equal in the two models. Informally we say that adding an orthogonal covariate doesn't change any of the estimated regression coefficients. Once we have found a $\mathbf{N}_{\text{ort}}^c(t-)$ being orthogonal to $\mathbf{Z}_1^{c,t}$, we can fit the model

$$dN_i(t) = \left\{ d\Theta_0(t) + d\Theta_1(t) Z_{i1}^{c,t} + d\Theta_2(t) N_{i,\text{ort}}^c(t-) \right\} Y_i(t) + dM_i(t), \qquad (13)$$

for individuals at risk at time t. The corresponding path diagram is Figure 8. Note that we use the same symbol $\Theta_1(t)$ for the regression function for $Z_{i1}^{c,t}$ as we did for the marginal model (11). The reason is that the regression function estimators for $\Theta_1(t)$ are identical in models (11) and (13), this being also the case for the two estimators for $\Theta_0(t)$.

In order to find an orthogonalized covariate as described above, it will suffice to orthogonalize $\mathbf{N}^c(t-)$ w.r.t. $\mathbf{Z}_1^{c,t}$. This can be done by fitting an ordinary linear least-squares regression of $\mathbf{N}^c(t-)$ on $\mathbf{Z}_1^{c,t}$ since such regression gives an orthogonal projection. Thus we start the derivation of the orthogonal covariate by fitting the standard linear model

$$\mathbf{N}^c(t-) = \Psi(t)\mathbf{Z}_1^{c,t} + \boldsymbol{\varepsilon}(t). \qquad (14)$$

The least-squares estimators are

$$\hat{\Psi}(t) = \frac{\sum N_i^c(t-) Z_{i1}^{c,t}}{\sum (Z_{i1}^{c,t})^2} \tag{15}$$

Defining a new covariate as the residual

$$\mathbf{N}_{\text{ort}}^c(t-) = \mathbf{N}^c(t-) - \hat{\Psi}(t)\mathbf{Z}_1^{c,t},$$

it is well known that this is orthogonal to $\mathbf{Z}_1^{c,t}$. The original dynamic covariate can be expressed as

$$\mathbf{N}^c(t-) = \hat{\Psi}(t)\mathbf{Z}_1^{c,t} + \mathbf{N}_{\text{ort}}^c(t-) \tag{16}$$

5.3 Estimation in the dynamic model

We want to investigate further the connection between the naive dynamic model (12) in Figure 7 and the orthogonal dynamic model (13) in Figure 8. By inserting (16) into (12), we discover that what we really are estimating in the naive dynamic model, is the model whose structural part is

$$dB_0(t) + dB_1(t)Z_{i1}^{c,t} + dB_2(t)\left\{\hat{\Psi}(t)Z_{i1}^{c,t} + N_{i,\text{ort}}^c(t-)\right\}. \tag{17}$$

Here we note that this expression contains the estimate $\hat{\Psi}(t)$, on which we are conditioning when fitting the additive model (in the same way that we are conditioning on the covariates, which is done in all regression analysis).

The structural part (17) can be re-written as

$$dB_0(t) + \left\{dB_1(t) + dB_2(t)\hat{\Psi}(t)\right\} Z_{i1}^{c,t} + dB_2(t)N_{i,\text{ort}}^c(t-), \tag{18}$$

which we recognize as the structural part of the orthogonal dynamic model (13).

By comparing the two expressions (13) and (18) for the structural part of the orthogonal dynamic model, we have found the connection between the two regression coefficients $dB_1(t)$ and $d\Theta_1(t)$ of the covariate $Z_{i1}^{c,t}$ for the naive dynamic model (12) of Figure 7 and the orthogonal dynamic model (13) of Figure 8. We get

$$d\Theta_1(t) = dB_1(t) + dB_2(t)\hat{\Psi}(t). \tag{19}$$

Here we note that $d\Theta_1(t)$ is the treatment effect in the marginal model. Thus we have shown that the treatment effect in the naive dynamic model is underestimated by $dB_2(t)\hat{\Psi}(t)$.

Based on the results above, we note that for the purpose of estimating the treatment effect $d\Theta_1(t)$ and the effect of the dynamic covariate $d\Theta_2(t)$, we don't actually have to perform any orthogonalization. Instead we can fit (12) to get $d\hat{B}_2(t)$, being equal to $d\hat{\Theta}_2(t)$, whereas the treatment effect can be found either from fitting the marginal model, or by fitting (12) and (14) and then calculate the right-hand side of (19).

5.4 Interpretation by path analysis

Up to now we have only used the path diagram as a visualizing tool of the regression of $dN(t)$ on covariates. However, Figure 9 shows a path diagram of the connection between all the variables, and it turns out that the estimation methods in path analysis is essentially identical to the orthogonalizing approach above, and gives in addition a better visual impression of what is going on.

The direct effect of the ancestor Z_1 on the descendant $dN(t)$ is represented by the path directly from Z_1 to $dN(t)$ and is denoted with a symbol next to the arrow in question. The symbol is defined as the regression coefficient of Z_1 when we fit the linear least-squares regression of $dN(t)$ on all its parents, in our case being Z_1 and $N(t-)$ (centered appropriately as described above). This means fitting (12), and then we have two of the three effects of Figure 9. For the remaining direct effect, we follow the definition of direct effect and fit the regression equation (14).

The indirect effect applies only to paths with intermediate nodes, and it is the product of all direct effects along subpaths of the path. Using Figure 9, the indirect effect of Z_1 on $dN(t)$ then is $\Psi(t) \cdot dB_2(t)$.

The total effect of Z_1 on $dN(t)$ is defined as the sum of all direct and indirect effects, and is then

$$\text{Total effect} = dB_1(t) + \Psi(t) \cdot dB_2(t), \tag{20}$$

thus from (19) the total effect has identical interpretation as the marginal treatment effect.

The result of this analysis is that the marginal model gives the total effect of treatment, which can presumably be identified with the causal effect. So what the is the point of the more complex analysis, couldn't one just be content with

performing a simple marginal analysis?

We believe the path analysis framework is illuminating. It connects the marginal model with the joint model and shows how the marginal effect of treatment can be decomposed into a direct and an indirect effect. Such a simple decomposition can only be performed simply because we postulate an additive, or linear, structure. In nonlinear structural models no simple relationship between marginal, direct and indirect effects exist. E.g. it would be difficult to perform a path analysis within a Cox proportional hazard framework.

The path analysis also shows the relationship between estimating treatment effects and dynamic effects. While the correct treatment effect can either be perceived as the regression effect in the marginal model or as the total effect in the joint model, the effect of the dynamic covariate is the corresponding regression effect in the joint model. Hence, the joint model and the associated path diagram is not unneccesary luxury, but required to get a correct estimation for both the treatment and the dynamic covariate.

5.5 The tumour data revisited

In Figure 2 it is seen that the effect of treatment in the marginal model is greater that the effect in the joint model. It is now clear that the total treament effect is found from the marginal model in the top right panel, while the direct effect of treatment in the full dynamic model is found in the right middle panel. The indirect effect of treatment is the difference between what is found in these two panels.

The direct effects of the dynamic covariates in the full model are found in the two bottom panels. We conclude that there is a clear effect of the number of previous occurences, but not of the time since the last event. Hence the process of repeated occurrences of tumours appears Markovian, but with different rates for different individuals.

There will be more uncertainty as how to interprete the results of the Cox analyses in Table 1. Clearly the effect of treatment would again be estimated from the marginal model, however the interpretation of the joint model is not clear since there is no simple connection between the joint and the marginal models as we have in the additive case. In particular, a definition of indirect

effects is lacking for the Cox model.

6 Variance estimators studied by simulation

The issue of variance estimation depends on whether the model is a true intensity model with martingale residuals, or just a rate model. Since the distinction between these two types of models is important here, we give a simulation study of how different variance estimators perform. The simulation is done similarly to Section 4.2, using the model $\lambda_T(t) = kZ_1 + (1-k)Z_2$ with $k = 0.25$, and where the Z's are independent and uniformly distributed on (0,1). A simulation study has been performed with each data set consisting of 40 individual processes on a time scale from 0 to 10. A total of $n = 200$ data sets have been simulated.

6.1 Conditional versus unconditional simulation

Note that we can simulate in two different ways with respect to the variables Z_{i1} and Z_{i2}. The conditional simulation approach is to simulate given Z_{i1} and Z_{i2}, meaning that we have a specific sample of individuals with fixed intensity, only simulating their processes given the intensity. The unconditional method is to simulate both Z_{i1} and Z_{i2} as well as the course of the processes at each repetition. The unconditional method is obviously the more general one of the two approaches since we are not restricting ourselves to a particular sample of individuals. In the conditional sampling scheme we will have only non-random covariates, thus simplifying the interpretation since one typically assumes, at least implicitly, non-random covariates in regression analysis. Unconditional simulation means that the variability of the results (for instance the variance estimator) has two components: one is the variability arising from the randomness at every time point of the process. The second variability component is that invoked by generating the covariate values (or generating the sample of individuals). The conditional simulation approach has only the first variability component. Hence the variance in the unconditional case will be larger than the variance in the conditional case, which also follows from the well known formula: $Var(Y) = Var(E(Y|X)) + E(var(Y|X)) > E(var(Y|X))$

The standard assumption in regression analysis is to assume fixed covari-

ates, that is the conditional case. Little theory exists on the more realistic case of random covariates, even for ordinary linear regression. In our case the distinction between the conditional and unconditional case turns out to be quite important.

6.2 Simulating the variances

We want to compare the martingale and Scheike variance estimators (Scheike, 2002) for the cumulative regression function estimates with each other, and also to compare them with the true variance. The true variance is not available analytically, but an estimate is the sample variance of the n simulated cumulative regression function estimates.

6.2.1 The variance of the total effect $\hat{\Theta}_1(t)$

A rather complex picture is emerging as regards the variances. We shall first consider estimation of the variance of the estimator of the fixed effect Z_1, distinguishing between conditional and unconditional simulation schemes as discussed above.

Let us first consider the marginal analysing model, that is, only using Z_1 in the analysis. Figure 10 (upper right panel) reveals that in the unconditional simulation scheme, the martingale variance of $\hat{\Theta}_1(t)$ is underestimating the variance in the marginal model, as we would expect from Section 2.2 since the martingale property isn't valid when one covariate is excluded. The Scheike variance estimator (Scheike 2002) , on the other hand, is seen to give a correct estimate, as expected.

It is interesting to note from the upper left panel of Figure 10 that when conditioning on Z_1 and Z_2, the martingale estimator is the correct one while the Scheike estimator overestimates the variance. Since the rates for each individual is in this case constant (not changing between simulations) the martingale property will be preserved whichever model is used.

In the lower right panel of Figure 10 we use a dynamic orthogonalized model. We see that both the Scheike and the martingale estimators underestimate the variance. It is reasonable that the estimators coincide closely since we have a full dynamic model where the martingale property would be expected to hold

approximately, cf. the results for residuals in Section 4.2. Compared with the marginal model discussed above, it is clear that the Scheike estimator will be expected to decrease when adding the dynamic covariate since the residuals will decrease due to the improved fit with two covariates, and Scheike's estimator is based on residuals. On the other hand the true variance will remain unchanged when adding the dynamic covariate since the orthogonality of the orthogonalized dynamic covariate to Z_1 makes the regression function estimators identical in the marginal model and the dynamic model.

The lower left panel of Figure 10 shows that both estimators are correct in the conditional scheme, this being reasonable since the model given Z_1 and the full model coincide as discussed above.

Regarding how to estimate the variance in a real-life situation, our main interest lies in the unconditional simulation scheme— since it reflects the total span of real life situations that we could face — and in the orthogonal dynamic model and the marginal model, being the models that we would use in a real life situation. As concluded above, in the marginal model the Scheike-estimator is the correct choice.

6.2.2 The variance of the dynamic effect $\hat{\Theta}_2(t)$

The variance of $\hat{\Theta}_2(t)$ appears by Figure 11 to be best estimated by the martingale variance estimator since it appears that the Scheike-estimator is underestimating the variance at the start. Once the processes have been going on for some time, the variance is in any case increasing only very slowly compared to the start. The reason for this is that the number of earlier events is a rather strong predictor on whether or not a new event will occur, but only when we have allowed the processes to run until a certain number of events have already happened.

7 Discussion

We have shown how to combine the analysis of a treatment effect and of dynamic covariates, and thereby getting a detailed insight into the structure of the data. Due to additivity this can be performed by path analysis, tying to-

gether marginal and joint models by defining direct, indirect and total effects. A similar analysis does not exist, for instance, for the Cox model where indirect effects have not been defined.

It should be noted that the biases studied here are different from those that are due to excluded or unknown confounders. For instance, in frailty theory it is well known that unknown risk factors will typically have the effect of attenuating the estimated effects of the observed ones. In other cases, excluding confounders may result in estimated effects that are too large. Had the unknown factors been observed and included in the analysis as confounders, then the estimated effects would be more correctly estimated. What we see here is the opposite phenomenon. The most correct estimates will be observed in marginal models, and including internal time-dependent covariates in the causal path between a fixed covariate (e.g. treatment) and the recurrent events may bias the estimate, often downwards. Hence, as pointed out repeatedly in the causal analysis literature, e.g. by Hernán et al (2002), whether to include or exclude covariates are crucially dependent on a causal understanding.

We believe the analysis given here clarifies issues which have been unsettled in the literature. For instance, Kalbfleisch and Prentice (2002, Section 9.4.3) study the effect of the medication thiotepa on the recurrence of superficial bladder tumours. In Table 9.3 of their book they present results for a number of different Cox models, including models with number of previous recurrences in some form. The effect of thiotepa is lower in the analyses which include this dynamic covariate just as is expected. The discussion by the authors is interesting, but does not seem conclusive with respect to which analysis one should choose. We believe an analysis along the lines of this paper would be illuminating.

It has been recognized over the last few years that in order to understand the issues of time-dependent covariates and time-dependent confounding, one has to look at causal models. This has forcefully been presented by Robins and coworkers, see e.g. Robins, Hernán and Brumback (2000). They apply the marginal structural models which introduce weighting procedures to counteract the biases introduced by confounding and selection. This is undoubtedly a generally valid procedure as long as one is mainly interested in correct estimation of, say, treatment effects. But if one wants to understand the more detailed

causal structure, then the marginal structural models may not be the right tool. Note that our path analysis is also a structural equation model, but different from that of Robins and coworkers.

Our approach can be extended to include several fixed covariates and several dynamic covariates as well as dynamic treatment regimes. As an example, consider the paper by Hernán, Brumback and Robins (2000). Here they study the impact on survival of the medication Zidovudine for HIV-positive patients. A simple analysis cannot be done because Zidovudine is being give to patients only when their CD4 count becomes quite low, hence when the disease is relatively advanced. Superficially, one might therefore get the impression that the treatment is associated with high risk of death, as indeed a simple Cox analysis shows. However, Hernán et al analyses the data according to the marginal structural model and finds, correctly, that the treatment reduces mortality. The complexity here is due to the fact that the CD4 count influences the probability that the treatment is started. On the other hand, treatment itself will influence the CD4 count. Hence, one should analyze not only how treatment influences survival, but also how it influences and is influenced by the CD4 count. It is natural to talk about the direct effect of treatment on mortality risk and the indirect effect working through the influence on the CD4 counts, and this distinction is just what path analysis offers. We believe our approach can also handle the example discussed by Hernán et al and be an alternative to the marginal structural model.

References

Aalen, O. O. (1980), A model for nonparametric regression analysis of counting processes, *in* 'Lecture Notes in Statistics, vol 2', Springer, pp. 1–25.

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.

Aalen, O. O., Fosen, J., Weedon-Fekjær, H., Borgan, Ø. & Husebye, E. (2004), 'Dynamic analysis of multivariate failure time data', *Biometrics* **60**, 764–773.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993), *Statistical models based on counting processes*, Springer.

Bollen, K. A. (1998), Path analysis, *in* P. Armitage & T. Colton, eds, 'Encyclopedia of Biostatistics', Wiley, pp. 3280–3284.

Eshima, N., Tabata, M. & Zhi, G. (2001), 'Path analysis with logistic regression models: Effect analysis of fully recursive causal systems of categorical variables', *J. Japan. Statist.* **31**, 1–14.

Gail, M. H., Santner, T. J. & Brown, C. C. (1980), 'An analysis of comparative carcinogenesis experiments based on multiple times to tumor', *Biometrics* **36**, 255–266.

Gandy, A. & Jensen, U. (2004), 'A nonparametric approach to software reliability', *Applied Stochastic Models in Business and Industry* **20**, 3–15.

Gruber, M. H. J. (1998), *Improving efficiency by shrinkage*, Marcel Dekker.

Hernán, M. Á., Brumback, B. & Robins, J. M. (2000), 'Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men', *Epidemiology* **11**, 561–570.

Hernán, M. Á., Brumback, B. & Robins, J. M. (2002), 'Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures', *Statistics in Medicine* **21**, 1689–1709.

Hernán, M. Á., Hernández-Díaz, S., Werler, M. M. & Mitchell, A. A. (2002), 'Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology', *American Journal of Epidemiology* **155**, 176–184.

Kalbfleisch, J. & Prentice, R. L. (2002), *The statistical analysis of failure time data*, 2 edn, Wiley.

Miloslavsky, M., Keleş, S. & van der Laan, M. J. (2004), 'Recurrent events analysis in the presence of time-dependent covariates and dependent censoring', *J. R. Statist. Soc. B* **66**, 239–257.

Pearl, J. (2000), *Causality: Models, reasoning and inference*, Cambridge University Press.

Pearl, J. (2001), Direct and indirect effects, *in* 'Proceedings of the Seventeenth Conference on Uncertainy in Artificial Intelligence, San Fransisco', Morgan Kaufmann, pp. 411–420.

Peña, E. A. & Hollander, M. (2004), Models for recurrent phenomena in survival analysis and reliability, *in* T. Mazzuchi, N. Singpurwalla & R. Soyer, eds, 'Mathematical Reliability: An Expository Perspective', Kluwer, pp. 105–123.

Robins, J. M., Hernán, M. Á. & Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology', *Epidemiology* **11**, 550–560.

Scheike, T. (2002), 'The additive nonparametric and semiparametric Aalen model as the rate function for a counting process', *Liftetime Data Analysis* **8**, 247–262.

Wright, S. (1934), 'The method of path coefficients', *Annals of Mathematical Statistics* **5**, 161–215.

Author's address

Postal address: Johan Fosen, University of Oslo, Department of statistics, Institute of Basic Medical Sciences, P.O. Box 1122 Blindern, N-0317 Oslo, Norway
Email: johan.fosen@basalmed.uio.no

Model	Covariate	Coef.	hazard ratio	se (coef.)	z-value	P-value
Marginal	Treatment	0.833	2.30	0.152	5.47	4.4e-8
Dynamic	Treatment	0.592	1.81	0.163	3.63	2.8e-4
	No. of prev. ev.	0.103	1.11	0.034	3.05	0.0023
	Time since last ev.	-0.320	0.73	0.169	-1.89	0.058

Table 1: Results of Cox analysis of the marginal and dynamic model on the mammary tumour data.

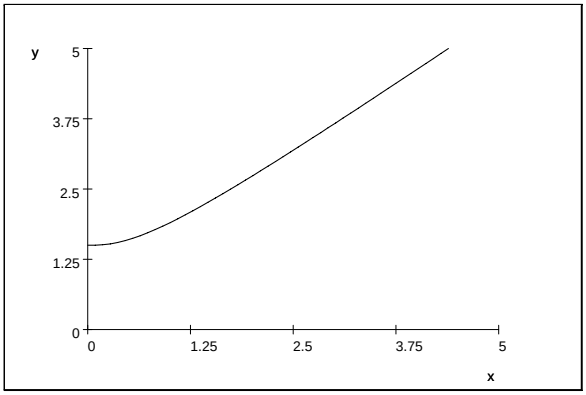

Figure 1: *Dependence on Z_1 in formula (7)*

	k									
Distribution	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$U(0,1)$	-23	-22	-17	-14	-9	-6	-4	-2	-1	0
$\exp(1)$	-65	-62	-56	-50	-42	-33	-22	-12	-2	0
$\Gamma(2,1)$	-54	-46	-43	-36	-28	-21	-13	-7	-2	0

Table 2: Relative decrease in percent of the cumulative regression function of Z_1 at the last time point when fitting the dynamic model instead of the marginal. Based on 1000 simulations of 40 individual processes generated from the model $\lambda(t) = 0.5 + kZ_1 + (1-k)Z_2$. Ridge regression with ridge factor 0.01 used in the presence of singularity.

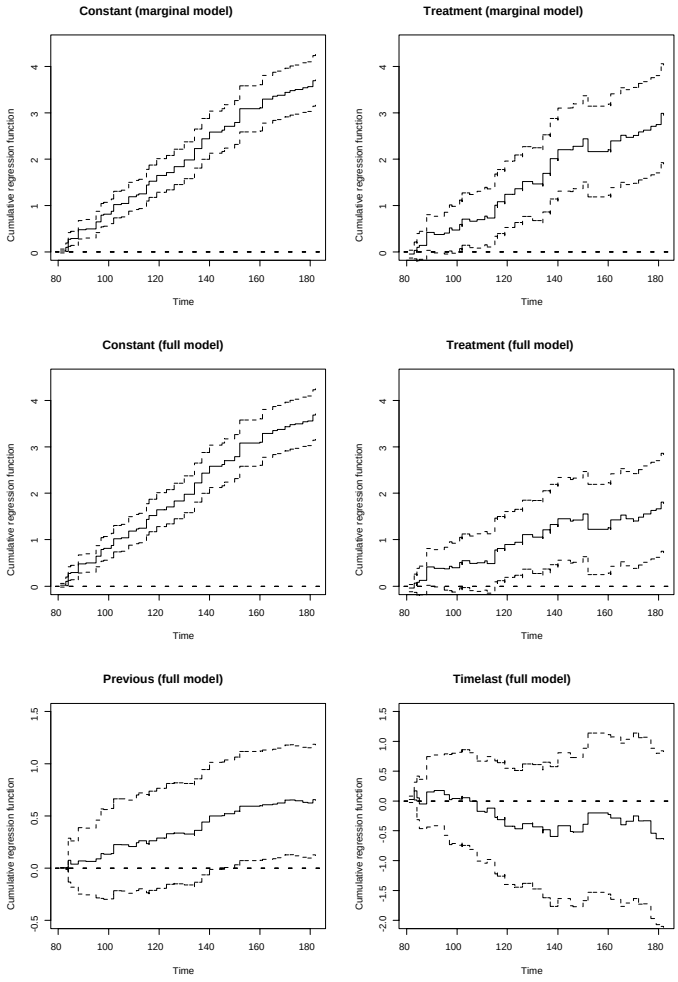

Figure 2: *Cumulative regression functions from the analysis of mammary tumour occurrences. The two upper figures are for the marginal model with only treatment included, whereas the four lower figures are for the model with all three covariates included. The two upper left figures are the baseline intensities, the two upper right are the cumulative regression function for treatment, the lower left is for the number of previous occurrences divided by observation time, the lower right is for time since previous event. Outer curves give pointwise 95% confidence intervals.*

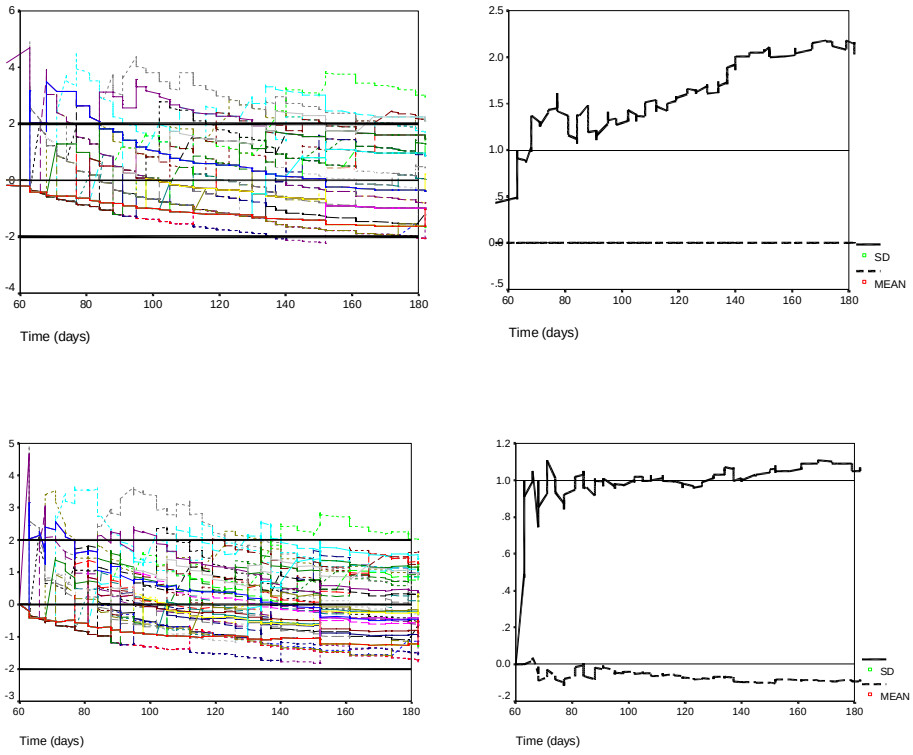

Figure 3: *Standardized residual processes (left panels) and mean and standard deviation of these processes (right panels) for the mammary tumour data. The results for the marginal and dynamic model are in the upper and lower panel respectively.*

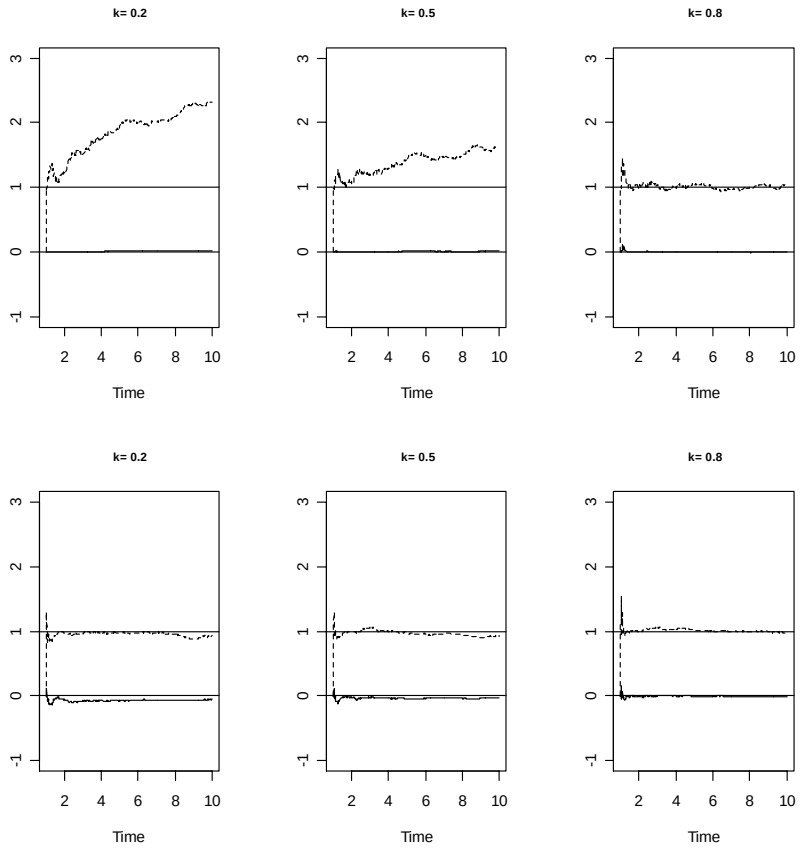

Figure 4: *Mean (solid curve) and standard deviation (dotted line) of the standardized martingale residuals of 40 processes. Shown for the marginal analyzing model (upper panels) and the dynamic analyzing model (the lower panels), and for different values of* k *when the true simulation model is* $\lambda(t) = 0.5 + kZ_1 + (1-k)Z_2$ *where the* Z*'s are gamma distributed with shape parameter 2 and scale parameter 1. Ridge regression with ridge factor 0.01 used in the presence of singularity.*

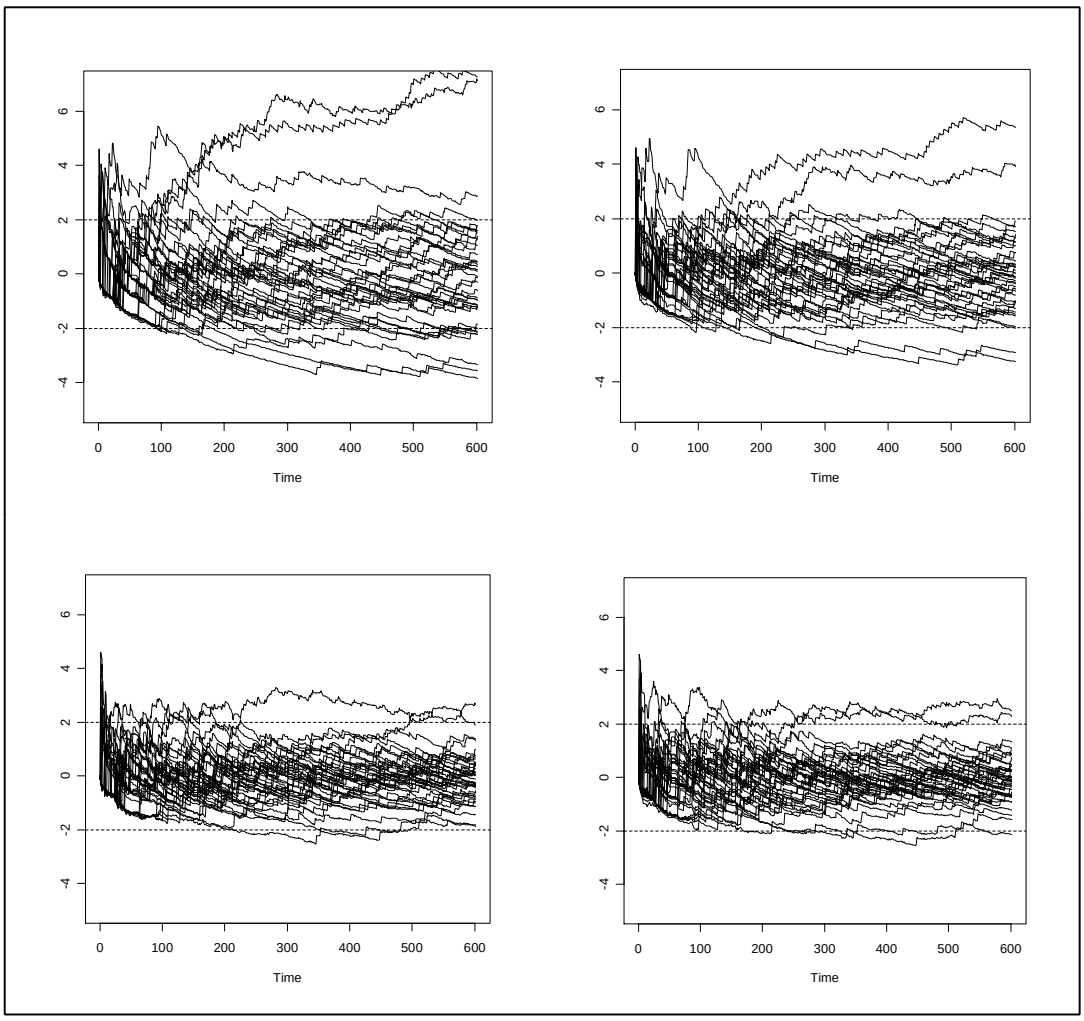

Figure 5: Standardized martingale residuals of 40 processes. Shown for the marginal analyzing model (upper panels) and the dynamic analyzing model (the lower panels), and for $k = 0.2$ (left) and $k = 0.5$ (right), where the true model is $\lambda(t) = 0.5 + kZ_1 + (1-k)Z_2$ with the Z gamma distributed with shape parameter 2 and scale parameter 1. Ridge regression with ridge factor 0.01 used in the presence of singularity.

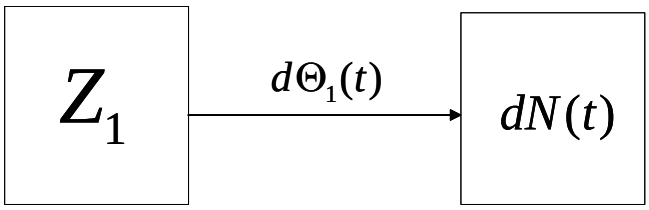

Figure 6: *Path diagram of the marginal model* (11).

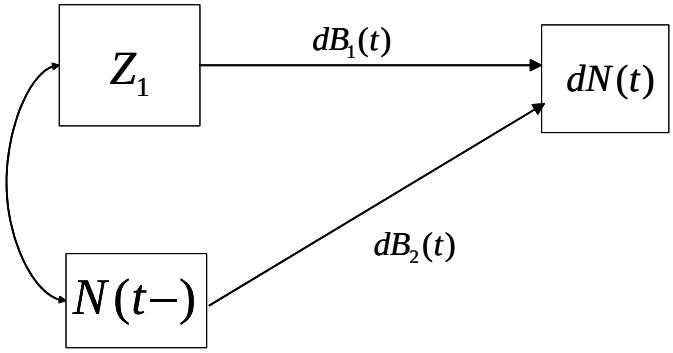

Figure 7: *Path diagram of the naive dynamic model* (12).

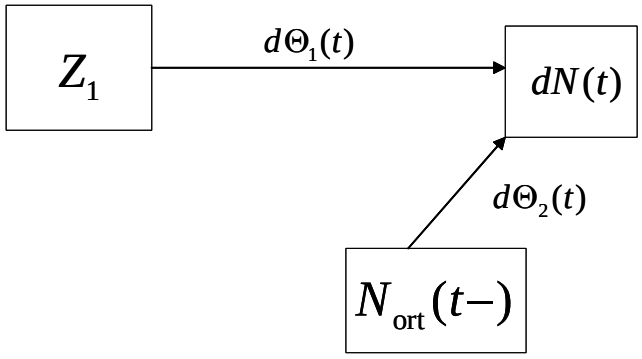

Figure 8: *Path diagram of the dynamic model (13) with orthogonal dynamic covariate.*

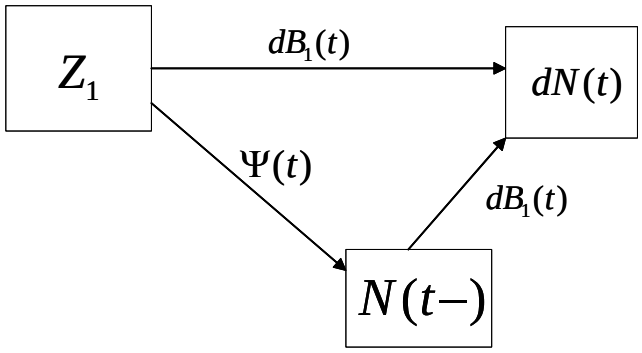

Figure 9: *Full path diagram of the dynamic model.*

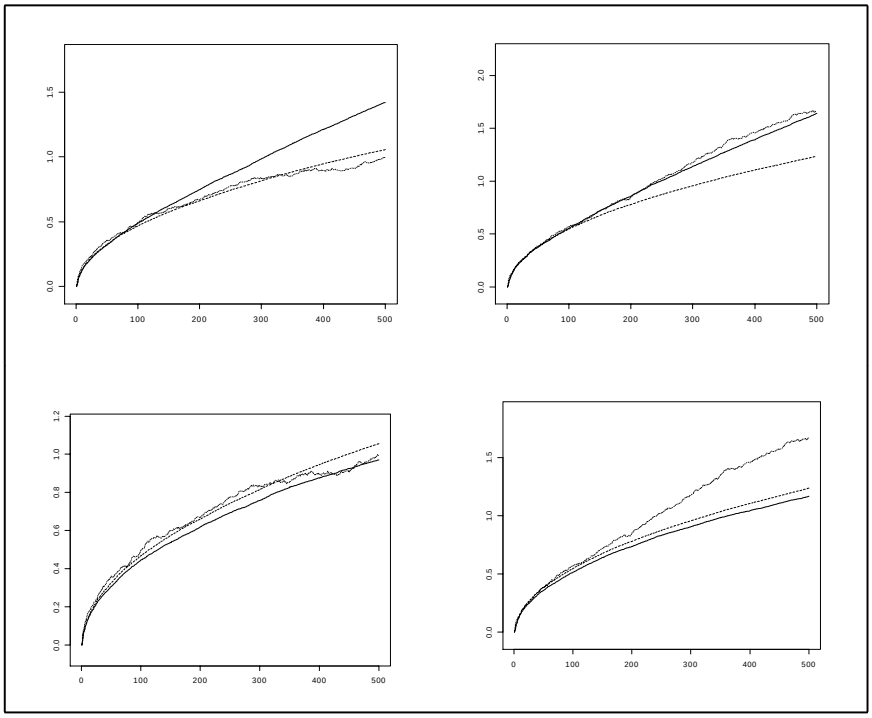

Figure 10: Variance-estimators for the cumulative regression function of covariate Z_1. Upper panel are marginal models and lower are dynamic. Conditional simulation scheme on the left and unconditional on the right. Solid line is the Scheike-estimator, dotted line is martingale-based estimator and dashed (irregular) line is the simulated variance.

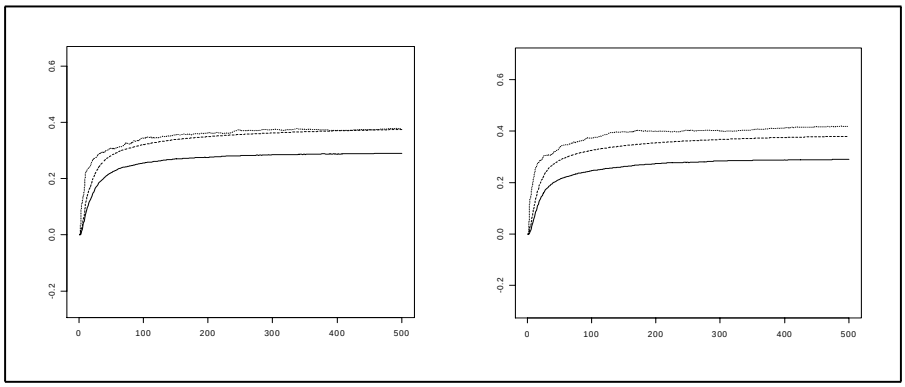

Figure 11: Variance-estimators for the cumulative regression function of covariate $N(t-)$ in the dynamic model. The left figure gives conditional simulation and the right unconditional. Solid line is the Scheike-estimator, dotted line is martingale-based estimator and dashed (irregular) line is the simulated variance.