Statistical Methods in Medical Research 2002; **11**: 91–115

Multi-state models for event history analysis

Per Kragh Andersen Department of Biostatistics, University of Copenhagen, Denmark and Danish Epidemiology Science Centre, Copenhagen, Denmark
Niels Keiding Department of Biostatistics, University of Copenhagen, Denmark

An introduction to event history analysis via multi-state models is given. Examples include the two-state model for survival analysis, the competing risks and illness–death models, and models for bone marrow transplantation. Statistical model specification via transition intensities and likelihood inference is introduced. Consequences of observational patterns are discussed, and a real example concerning mortality and bleeding episodes in a liver cirrhosis trial is discussed.

1 Introduction

Event history analysis deals with data obtained by observing individuals over time, focusing on events occurring for the individuals. Thus, typical outcome data consist of *times of occurrence* of events and the *types of events* that occur. Frequently, an event may be considered as a *transition* from one state to another and, therefore, *multi-state models* will often provide a relevant modeling framework for event history data. Multi-state models are discussed from several points of view in the books by Andersen *et al.*,[1] Blossfeld and Rohwer,[2] Courgeau and Lelièvre[3] and Hougaard[4] (Chapters 5 and 6); see Hougaard[5] and Commenges[6] for recent survey papers.

The purpose of the present article is to provide an overview of the topic and to serve as an introduction to the other articles in this issue.

2 Survival data

The most simple multi-state model is the two-state model for survival data with one transient state '0: alive' and one absorbing state '1: dead' (see Figure 1). In general, an *absorbing* state is a state from which further transitions cannot occur while a *transient* state is a state that is not absorbing. The observation for a given individual will here in the most simple form consist of a random variable, say T, representing the time from a given origin (time 0) to the occurrence of the event 'death'. The distribution of T may be characterized by the probability distribution function $F(t) = \text{Prob}(T \leq t)$ or, equivalently, by the survival distribution function $S(t) = 1 - F(t) = \text{Prob}(T > t)$. It is seen that $S(t)$ and $F(t)$, respectively, correspond to the probabilities of being in state 0 or 1 at time t. If every individual is assumed to be in state 0 at time 0 then $F(t)$ is also the

Address for correspondence: Department of Biostatistics, University of Copenhagen, Blegdamsvej 3, DK 2200 Copenhagen N, Denmark. E-mail: pka@biostat.ku.dk

 10.1191/0962280202SM276ra

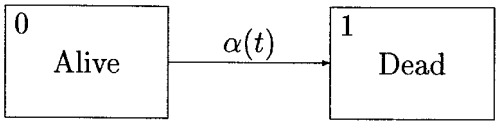

Figure 1 The two-state model for survival data.

transition probability from state 0 to state 1 for the time interval from 0 to t. In continuous time the distribution of T may also be characterized by the *hazard rate function*

$$\alpha(t) = -\mathrm{d}\log S(t)/\mathrm{d}t = \lim_{\Delta t \to 0} \frac{\mathrm{Prob}(T \leq t + \Delta t | T \geq t)}{\Delta t}$$

that is,

$$S(t) = \exp\left(-\int_0^t \alpha(u)\mathrm{d}u\right)$$

Thus, $\alpha(\cdot)$ is the *transition intensity* from state 0 to state 1, i.e., the instantaneous probability per time unit of going from state 0 to state 1.

In general, event history analysis deals with inference for transition intensities and transition probabilities in multi-state models. This includes estimation and hypothesis tests for these quantities and analysis of regression models where these quantities are related to (possibly time-dependent) explanatory variables observed for the individuals under study. Most frequently, multi-state models are defined by their transition intensities from which transition probabilities may or may not be derived depending on the modeling assumptions. This latter activity is sometimes denoted *survival synthesis*.

A typical feature of event history analysis is the inability to observe complete event histories, for example by the end of the observation period all individuals under study may not have reached an absorbing state. In survival analysis this would correspond to individuals still being alive by the end of the study and this kind of *incomplete observation* is known as *right-censoring*. Furthermore, all individuals may not have been observed from the same time origin. This kind of incomplete observation where individuals are only observed conditionally on not having reached an absorbing state by the time of initiation of the study is known as *left-truncation*. Restricting attention to right-censoring a crucial problem is whether the available incomplete data enables one to make valid inference on parameters in the multi-state model for the complete data. The condition for this is known as *independent* right-censoring and the interpretation is that a sample observed after independent right-censoring is 'representative' for the population without censoring. This means that individuals who are censored should have neither lower nor higher risk of future events than individuals who are not censored, see for example, Andersen *et al.*,[1] Chapter III and Kalbfleisch and Prentice, p. 120.[7]

3 Multi-state models

In this section we will present a number of different multi-state models. We will begin with a few heuristic definitions, which may all be made rigorous within the framework of so-called *marked point processes*.[1,8]

A *multi-state process* is a stochastic process $(X(t),\ t \in \mathcal{T})$ with a finite *state space* $\mathcal{S} = \{1, \ldots, p\}$ and with right-continuous sample paths: $X(t+) = X(t)$. Here, $\mathcal{T} = [0,\ \tau]$ or $[0,\ \tau)$ with $\tau \leq +\infty$. The process has *initial distribution* $\pi_h(0) = \text{Prob}(X(0) = h)$, $h \in \mathcal{S}$. A multi-state process $X(\cdot)$ generates a *history* $\mathcal{X}_t$ (a σ-algebra) consisting of the observation of the process in the interval $[0,\ t]$. Relative to this history we may define *transition probabilities* by:

$$P_{hj}(s,\ t) = \text{Prob}(X(t) = j | X(s) = h, \mathcal{X}_{s^-})$$

for $h, j \in \mathcal{S}, s, t \in \mathcal{T}, s \leq t$ and *transition intensities* by the derivatives

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} \frac{P_{hj}(t,\ t + \Delta t)}{\Delta t}$$

which we shall assume exist. Some transition intensities may be 0 for all t. Graphically, multi-state models may be illustrated using diagrams with boxes representing the states and with arrows between the states representing the possible transitions, i.e., the non-zero transition intensities.[1,9] A state $h \in \mathcal{S}$ is *absorbing* if for all $t \in \mathcal{T}, j \in \mathcal{S}$, $j \neq h, \alpha_{hj}(t) = 0$; otherwise h is *transient*. The *state probabilities* $\pi_h(t) = \text{Prob}(X(t) = h)$ are given by:

$$\pi_h(t) = \sum_{j \in \mathcal{S}} \pi^j(0) P_{jh}(0,\ t)$$

Notice that the $P_{hj}(\cdot,\ \cdot)$ and thereby the $\alpha_{hj}(\cdot)$ depend on both the probability measure Prob and on the history, though this dependence has been suppressed in the notation. If $\alpha_{hj}(t)$ only depends on the history via the state $h = X(t)$ occupied at t then the process is Markovian. Sometimes one is interested in considering an extended history which also includes observed *covariates*. If only time-fixed covariates Z are studied then the observed history is $\mathcal{F}_t = \mathcal{X}_t \vee \mathcal{Z}_0$ whereas time-dependent covariates $Z(t)$ give rise to an extended history of the form $\mathcal{F}_t = \mathcal{X}_t \vee \mathcal{Z}_t$ where in both cases $\mathcal{Z}_t$ is the history generated by the covariates in $[0,\ t]$. (Here, for σ-algebras $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \vee \mathcal{B}$ is the smallest σ-algebra containing both $\mathcal{A}$ and $\mathcal{B}$). We shall here focus on the *purely endogeneous* case where $\mathcal{Z}_t \subset \mathcal{X}_t \vee \mathcal{Z}_0$, i.e., the covariates are either all time-fixed or the random development of the time-dependent covariates is fully specified by the history of the process itself (see, however, Section 5.4 for cases with time-dependent covariates that are not endogeneous).

3.1 The two-state model for survival data

This model, illustrated in Figure 1, has $p = 2$ states and only one possible transition from state 0 to 1. The corresponding transition intensity $\alpha_{01}(t)$ is given by the hazard rate function $\alpha(t)$, while $\alpha_{10}(t) = 0$ for all t, that is, state 1 is absorbing. The initial distribution is degenerate in 0: $\pi_0(0) = 1$ and the process is *Markovian*. Covariates may be entered into the model using a regression model for $\alpha(\cdot)$. Examples of this simple model are, of course, numerous and will not be considered here.

3.2 The competing risks model

This model has one transient state '0: alive' and a number, k, of absorbing states, state h, $h = 1, \ldots, k$ corresponding to 'death from cause h'. Thus, there are $p = k + 1$ states. The model is illustrated for $k = 2$ in Figure 2.

The transition intensities $\alpha_{0h}(t)$ for $h = 1, \ldots, k$ are given by the cause-specific hazard functions:

$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{\text{Prob}(\textit{Dead from cause } h \textit{ by } t + \Delta t | T \geq t)}{\Delta t}$$

where T is the survival time. The initial distribution is degenerate in 0, the only transient state of the model, i.e., $\alpha_{hj}(t) = 0$ for all $h \neq 0$ and all j. The transition probabilities are given by the survival function

$$P_{00}(0,\ t) = S(t) = \text{Prob}(T > t) = \exp\left(- \int_0^t \sum_{h=1}^{k} \alpha_h(u) \mathrm{d}u \right)$$

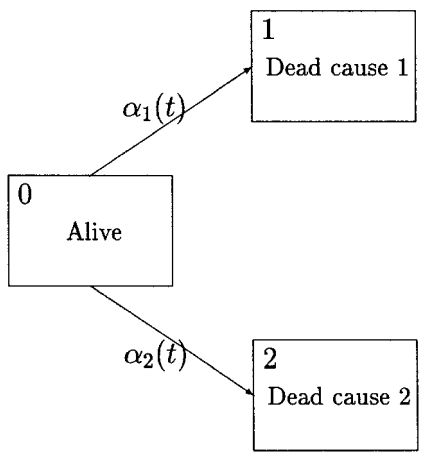

Figure 2 Competing risks model for mortality from two causes.

and the cumulative incidence functions

$$P_{0h}(0,\ t) = \int_0^t S(u-)\alpha_h(u)\mathrm{d}u, \qquad h = 1, \ldots, k$$

Like the simple two-state model ($k = 1$) the competing risks model is Markovian and covariates may be included into the model via regression models for the cause-specific hazards. This model, with examples, will be studied in detail in the companion paper.[10]

3.3 The illness–death model

This model is illustrated in Figure 3. Often the time t is the age of the individual, and usually individuals will be assumed to be in state 0 at $t = 0$. However, individuals will not always be observed from $t = 0$ as shall be further discussed in Sections 4 and 5. The mortality $\alpha_{12}(t)$ of the diseased (the *lethality*) may sometimes depend on duration d since entry to state 1 in addition to the dependence on 'age' t. (Notice that, despite the notation, $\alpha_{12}(t)$ then depends on the *random* time of the most recent transition into 1). If $\alpha_{12}(t)$ does not depend on d the process is Markovian, otherwise it is a *semi-Markov process*, an example of a purely endogeneous process.

In Figure 3 we have indicated the possibility of *reversibility*: the transition back from state 1 to 0 is possible. It will turn out that the simple unidirectional model in Figure 4 is rather easier to analyse statistically. Thus the transition probabilities in this model have simple explicit expressions:

$$P_{00}(s,\ t) = \exp\left(-\int_s^t (\alpha_{02}(u) + \alpha_{01}(u))\mathrm{d}u\right)$$

and (in the Markovian case)

$$P_{01}(s,\ t) = \left(\int_s^t P_{00}(s,\ u-)\alpha_{01}(u)P_{11}(u,\ t)\mathrm{d}u\right) \tag{1}$$

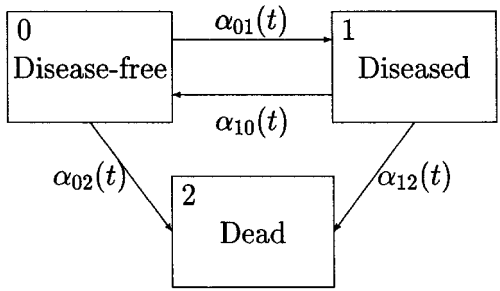

Figure 3 The illness–death or disability model.

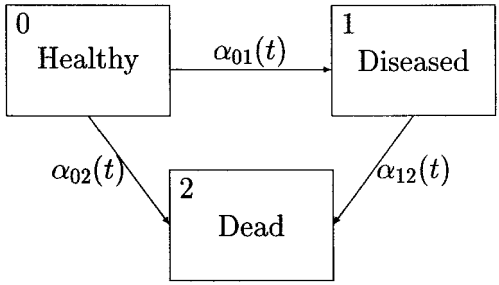

Figure 4 Unidirectional illness–death model.

where

$$P_{11}(s,\ t) = \exp\left(-\int_s^t \alpha_{12}(u)\mathrm{d}u\right)$$

More generally, the lethality $\alpha_{12}(\cdot)$ may depend on both age and duration. If we then define

$$\alpha_{12}(t,\ d) = \lim_{\Delta t \to 0} \frac{\mathrm{Prob}(X(t+\Delta t) = 2 | X(t) = 1, 0 \to 1 \text{ transition at } t-d)}{\Delta t},$$

$P_{11}(u,\ t)$ in (1) should be replaced by $\exp(-\int_u^t \alpha_{12}(s,\ s-u)\mathrm{d}s)$. The illness–death model is one of the most important multi-state models and it was discussed in early papers by Fix and Neyman[11] and Sverdrup.[12]

An example of the application of such a model is provided by the PROVA study group.[13] In brief, PROVA was a Danish multicenter clinical trial with the purpose of evaluating the effect of propranolol and/or sclerotherapy versus no treatment on bleeding and survival in patients with liver cirrhosis. Eligible patients included those in whom cirrhosis was histologically verified and where endoscopy had shown oesophageal varices, but who had not yet experienced a transfusion-requiring bleeding from the varices. Between 1985 and 1989, 286 patients were randomized as summarized in Table 1, which also shows the number of events observed before 1 January 1990, i.e., the observed numbers of transitions in the multi-state model. We shall return to this example in Section 5.5 below.

Table 1 Treatment allocation and number of end-points observed in the PROVA trial

Treatment group	Patients	Bleedings $0 \to 1$	Deaths without bleeding $0 \to 2$	Deaths after variceal bleeding $1 \to 2$
Sclerotherapy only	73	13	13	5
Propranolol only	68	12	5	6
Both treatments	73	12	20	10
No treatment (control)	72	13	8	8
Total	286	50	46	29

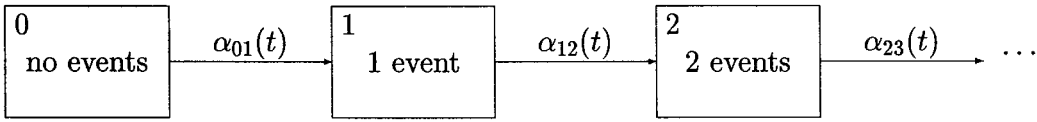

Figure 5 A model for repeated events.

3.4 Repeated events

If interest focuses on repeated occurrences of a given event, for example, hospital admissions, childbirths, infections etc., then a model as illustrated in Figure 5 (where transitions to an absorbing 'Dead' state have been omitted) may be considered. In applications of such a model an interesting functional is often the expected number of occurrences of the event over the time interval $[0, t]$.[14]

3.5 Interaction between life history events

This Markov model, illustrated in Figure 6, describes the joint behaviour of two life events A and B; if $\alpha_{0B} = \alpha_{A,AB}$ but $\alpha_{0A} \neq \alpha_{B,AB}$, A is called *locally dependent on B* but B is not locally dependent on A. The temporal order of events allows for this *asymmetric* concept of dependence, which yields more information for drawing causal inference than the standard symmetric association concepts from cross-sectional studies. Similar duration dependance as in the illness–death process might be added. A model of this type has been discussed for a study of interaction between menopause and a certain chronic skin disease.[2,3,15]

3.6 Bone marrow transplantation

A model combining most of the above features has been studied in detail as describing some of the possible states of a leukaemia patient following bone marrow transplantation (see Figure 7).[16,17]

Patients have been given various kinds of therapy to temporarily keep the disease down; they are said to be in *remission*. In our context these patients are followed from bone marrow transplantation ($t = 0$), initially considered in state 0. Two different types

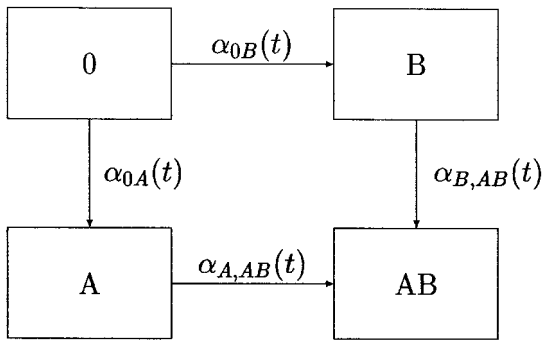

Figure 6 Interaction between life history events.

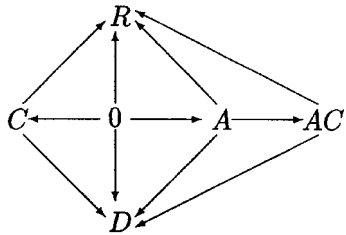

Figure 7 A model for events following bone marrow transplantation: acute and chronic graft-versus-host disease, relapse and death.

of complications are considered: acute graft-versus-host disease (A), chronic graft-versus-host disease (C), and a special state AC is defined for those patients acquiring both A and C. Patients are followed until relapse of the leukaemia (R) or death (D) while still in remission. Relapsed patients are not followed further in this context. If all transition rates depend only on time t since transplantation, we have again a Markov process, but various kinds of duration dependence (semi-Markov process models) may also be relevant. A model of this type will be discussed in more detail in the companion paper by Klein and Shu.[18]

4 Counting process representation, likelihood

Assume that multi-state processes $X_i(t)$ such as those described in Section 3 are observed over intervals $[0,\ \tau_i]$ for individuals $i = 1, \ldots, n$. Assume first that τ_i is a fixed (i.e., non-random) time of termination of observation for individual i. Random right-censoring (Section 2) and delayed entry are treated below. Since $X_i(t)$ is constant between transitions, it is equivalent to record $X_i(0)$ and the *counting processes*

$$N_{hj}^i(t) = \#\ (\text{direct transitions } h \to j \text{ in } [0,\ t] \text{ for } i);$$

described by the times T_{hj}^{ik} of these transitions, where

$$0 < T_{hj}^{i1} < \cdots < T_{hj}^{iN_{hj}^i(\tau_i)} \leq \tau_i$$

Let $N_{hj}(t) = \sum_{i=1}^n N_{hj}^i(t)$. It will also be useful to introduce $Y_h^i(t) = I\{X_i(t-) = h\}$ and

$$Y_h(t) = \#\ (\text{individuals 'at risk' in state } h \text{ at time } t-) = \sum_{i=1}^n Y_h^i(t)$$

Note that, for $t > \tau_i$, $N_{hj}^i(t) = N_{hj}^i(\tau_i)$ and $Y_h^i(t) = 0$; these can thus be considered to be defined on $(0,\ \infty)$.

For individual i denote the initial distribution $(\pi_h^i(0))$, the density of time-fixed covariates $f(Z_i)$, and the transition intensities $\alpha_{hj}^i(t)$, then the *likelihood* is[1]

$$\prod_{i=1}^{n} f(Z_i)\pi_{X_i(0)} \prod_{h\neq j} \prod_{k=1}^{N_{hj}^i(\tau_i)} \alpha_{hj}^i(T_{hj}^{ik}) \exp\left(-\int_0^{\tau_i} \alpha_{hj}^i(t) Y_h^i(t)\, \mathrm{d}t\right) \tag{2}$$

It is very common to condition on Z_i and on the initial values $X_i(0)$ (the distribution of which may often be degenerate anyway), and consequently omit the factors $f(Z_i)\pi_{X_i(0)}$ from the likelihood. We shall do so without further comment in the sequel.

Recall from above that the notation $\alpha_{hj}^i(t)$ represents possible dependence of the transition intensity on the whole history $\mathcal{X}_t^i$ of the process. Thus, $\alpha_{hj}^i(t)$ may well contain covariates and other random elements, as already exemplified.

Two patterns of incomplete observations are particularly easily tractable, because they lead to only minor modification of this likelihood: *delayed entry*, where individual i enters at some time V_i; and *right-censoring* where nothing is known about i after some time U_i. Both V_i and U_i may be random although either only dependent on the previous history of the process or independent of the process (see for example Andersen *et al.*,[1] Chapter 3, for precise specification of this and further discussion). The reason for the particular tractability of these mechanisms is that the 'at risk' indicator $Y_h^i(t) = I\{X_i(t-) = h\}$ in the likelihood just needs to be amended to

$$Y_h^i(t) = I\{X_i(t-) = h, \qquad V_i < t \leq U_i\}.$$

5 Statistical model specification

As indicated in the introduction the first purpose of event history analysis is to gain insight into the dynamics of the processes by quantifying *transition intensities* and perhaps assessing their dependence on covariates, possibly using various stratifications. Sometimes additional functionals are useful, particularly various types of *transition probabilities* obtained by integrating certain functions of the transition intensities. A final purpose may be *prediction*, both as illustration of the dynamics and for concrete practical purposes.

5.1 Markov processes

The most important class of models is the (continuous time) *Markov process* $X(t)$ on the finite state space $\mathcal{S} = \{1, \ldots, p\}$ where the dependence of $\alpha_{hj}^i(t)$ on the history $\mathcal{X}_t$ introduced at the beginning of Section 3 is only via the current state of $X(t)$ (and possibly via time-fixed covariates). Statistical models are usually obtained by specifying the class of transition intensities $(\alpha_{hj}^i(t))$ for each individual i.

5.1.1 Constant and piecewise constant transition intensities

The simplest class of models is obtained by keeping the transition rates constant: $\alpha^i_{hj}(t) = \alpha^i_{hj}$. *Piecewise constant* intensities

$$\alpha^i_{hj}(t) = \alpha^{i(l)}_{hj}, \qquad t^{hj}_{l-1} < t \leq t^{hj}_l, \qquad \text{all } t_0 = 0$$

form the next step up and this choice is of widespread use, particularly in large studies in econometrics, epidemiology, sociology and demography.[1,9,19,20]

Assume first that all individuals have the same transition intensities, $\alpha^{i(l)}_{hj} = \alpha^{(l)}_{hj}$. The likelihood (2) then simplifies to

$$\prod_l \prod_{h \neq j} (\alpha^{(l)}_{hj})^{N^{(l)}_{hj}} \, \mathrm{e}^{-\alpha^{(l)}_{hj} S^{(l)}_h}$$

where $N^{(l)}_{hj} = N_{hj}(t^{hj}_l) - N_{hj}(t^{hj}_{l-1})$ and

$$S^{(l)}_h = \sum_i \int_{t^{hj}_{l-1}}^{t^{hj}_l} Y^i_h(t) \, \mathrm{d}t$$

Since this likelihood looks like one resulting from observing $N^{(l)}_{hj}$ events in a Poisson process with intensity $\alpha^{(l)}_{hj}$ observed over the interval $(0, S^{(l)}_h)$, except that here $S^{(l)}_h$ is random, this likelihood, and even the model, are often associated with Poisson's name. Maximum likelihood estimation is elementary, yielding

$$\hat{\alpha}^{(l)}_{hj} = \frac{N^{(l)}_{hj}}{S^{(l)}_h}$$

the classical *occurrence/exposure rate*. Asymptotic inference may be obtained from the observed information

$$-D^2 \log L = \frac{N^{(l)}_{hj}}{(\alpha^{(l)}_{hj})^2}$$

yielding variance estimates

$$\mathrm{Var}(\hat{\alpha}^{(l)}_{hj}) \sim \frac{(\alpha^{(l)}_{hj})^2}{N^{(l)}_{hj}} \sim \frac{N^{(l)}_{hj}}{(S^{(l)}_h)^2}$$

and all estimators asymptotically independent.

Transition probabilities for the constant and piecewise constant Markov process models are explicit functions of the transition intensities,[21] allowing direct 'plug-in' maximum likelihood estimation, as well as calculation of standard error estimates via the delta method.

5.1.2 Other parametric models for the transition intensities.

Although the piecewise constant model is often sufficient to describe the dependence of intensities on time, other possibilities exist. Certain mathematical functions of time may generate the model, such as the Gompertz–Makeham model for mortality

$$\alpha(t) = \alpha + \beta\gamma^t$$

but except for mortality studies in actuarial and some demographic contexts such parametric models are little used. One reason for this may be the powerful development of methodology for 'nonparametric' statistical inference, where $\alpha_{hj}(t)$ is left unspecified.

5.1.3 Freely varying ('nonparametric') transition intensities.

Assume first that the transition intensities are the same for all individuals but that they are allowed to vary freely with time: $\alpha^i_{hj}(t) = \alpha_{hj}(t)$. Statistical inference is then conveniently phrased in terms of the counting process approach pioneered by Aalen[22,23] (see Andersen *et al.*[1] for a detailed exposition). Estimators (which may be given a nonparametric maximum likelihood interpretation) of the integrated intensities

$$A_{hj}(t) = \int_0^t \alpha_{hj}(u)\,\mathrm{d}u$$

are obtained as the *Nelson–Aalen estimators*

$$\hat{A}_{hj}(t) = \int_0^t \frac{J_h(u)}{Y_h(u)}\,\mathrm{d}N_{hj}(u) = \sum_i \sum_{k:0<T^{ik}_{hj}<t} \frac{1}{Y_h(T^{ik}_{hj})} \tag{3}$$

$J_h(u) = I\{Y_h(u) > 0\}$, with variance estimators

$$\hat{\sigma}^2(\hat{A}_{hj}(t)) = \int_0^t \frac{J_h(u)}{Y_h(u)^2}\,\mathrm{d}N_{hj}(u) = \sum_i \sum_{k:0<T^{ik}_{hj}<t} \frac{1}{Y_h(T^{ik}_{hj})^2}$$

An elaborate mathematical theory based on stochastic integrals and martingales is available to study exact and asymptotic properties of these estimators.

When estimates are desired of the transition intensities $\alpha_{hj}(t)$ themselves rather than their integrals, *smoothing* techniques are necessary.[1]

An important feature of the nonparametric approach is its elegant generalization,[24] to estimating *transition probabilities*. The basic tool is the (matrix) *product integral*. Let I be the identity matrix and G a matrix-valued function. The corresponding product integral is defined as

$$\Pi_0^t(I + G(ds)) = \lim_{\max|t_\nu - t_{\nu-1}|\to 0} \prod (I + G(t_\nu) - G(t_{\nu-1}))$$

where $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of $[0, t]$. In particular, if G is continuous and scalar

$$\Pi_0^t(1 + G(ds)) = e^{G(t)-G(0)}$$

and if G is a scalar step function

$$\Pi_0^t(1 + G(ds)) = \prod_{k=1}^{K}(1 + \Delta G(t_{(k)}))$$

where $t_{(0)} = 0$ and $0 < t_{(1)} < \cdots < t_{(K)} \leq t$ are the jump times of G and

$$\Delta G(t_{(k)}) = G(t_{(k)}) - G(t_{(k-1)})$$

Define $\alpha_{hh}(t) = -\sum_{j \neq h} \alpha_{hj}(t)$ and the intensity matrix function $A(t) = (\alpha_{hj}(t))$; then the matrix $P(s, t) = (P_{hj}(s, t))$ of transition probabilities

$$P_{hj}(s, t) = \text{Prob}(X_i(t) = j | X_i(s) = h)$$

is given by

$$P(s, t) = \Pi_s^t(I + A(du))$$

The *Aalen–Johansen* estimator of $P(s, t)$ is obtained by plugging the matrix of Nelson–Aalen estimators $(\hat{A}_{hj}(t))$ into the formula:

$$\hat{P}(s, t) = \Pi_s^t(I + \hat{A}(du))$$

For the simple two-state model for survival data $\hat{P}_{00}(0, t)$ reduces to the classical Kaplan–Meier[25] estimator $\hat{S}(t) = \prod_{T_i \leq t}(1 - dN_{01}(T_i)/Y_0(T_i))$ of the survival function $S(t)$.

As documented in detail,[1] there is a well-developed theory, again based on stochastic integrals and martingales, about the asymptotic properties of the Aalen–Johansen estimator.

5.1.4 Markov regression models.

For Markov models with several states, there will often be too little empirical basis for estimating freely varying transition intensities between all states for all subgroups, so that more parsimonious regression models are required. The most frequently used regression models in event history analysis have a multiplicative structure with a baseline $h \to j$ transition intensity $\alpha_{hj0}(t)$, assumed common for all individuals. For an individual, i, with time-fixed covariates $Z_i = (Z_{im})$ the transition intensity is then modelled as

$$\alpha_{hj}^i(t) = \alpha_{hj0}(t) \exp(\beta_{hj}' Z_i) \tag{4}$$

where the effect of a covariate Z_{im} is described by factors of proportionality $\exp(\beta_{hjm})$. In (4) the baseline hazard may be completely unspecified as in the Cox proportional hazards model for survival data or it may be assumed to be piecewise constant leading to Poisson regression models.[26,27] In both cases, inference may be based on the likelihood (2), which for the Cox model leads to the so-called Cox's partial likelihood.[1,28] The choice between Cox and Poisson models is frequently a matter of convenience, though the latter may be advantageous in large studies where a sufficiency reduction of data into tables of event counts and person-years within groups of (categorical) covariates is feasible.[19] In contrast (Section 5.6) application of the Cox model requires one data record per individual for each transition.

In (4) the notation suggests that separate baseline hazards and regression coefficients are assumed for each possible transition. If that is the case then the parameters may be estimated by fitting separate Cox or Poisson models for each transition. However, more parsimonious models may be obtained by assuming some baseline transition intensities proportional[16,29] or by assuming some covariates to have the same effect on several transitions.[1] Also, models where the proportional hazards assumption is relaxed may be considered. In the Poisson case this is simply an interaction between time and the covariate giving rise to non-proportionality whereas, for the Cox model, the less restrictive model is known as the *stratified* Cox model.

In Section 5.6 we describe how such flexible Cox models may be formulated in a way that shows how standard computer software may be applied.[1] In a similar way, Poisson regression models may be analysed using standard generalized linear models software.[30,31]

In Section 5.6 we also briefly discuss how to perform 'survival synthesis', i.e., to combine the regression estimates for the transition intensities into transition or state probability estimates.

Another regression model for survival data that readily extends to multi-state models is Aalen's non-parametric additive model:[1,32,33]

$$\alpha_{hji}(t) = \alpha_{hj0}(t) + \beta'_{hj}(t)Z_i$$

In this model both the baseline transition intensities $\alpha_{hj0}(t)$ and the regression functions $\beta_{hjm}(t)$ are left unspecified and non-parametric estimates may be obtained using a generalized least squares procedure. Aalen *et al.* presented a review of this model and its use in multi-state models.[34]

5.2 Beyond Markov processes

The most important deviations from the Markov property in practice are various kinds of *duration dependence*, where transition intensities depend on other time origins than $t = 0$, typically the time at entry to the present state. There are two main approaches to handling these.

As long as transition intensities depend only on one time origin each (for example, all intensities depend only on duration in the present state), a model for the multistate process may be obtained by combining independent submodels for each transition intensity. These may, in turn, be modelled as constant or piecewise constant or by non-

or semiparametric models, and as long as there is a unidirectional flow in the model, transition probabilities are still straightforward explicit functionals, which may be estimated by plugging in the intensity estimates. Variance calculations may however become less direct.

More elaborate models will include several time origins (such as age, disease duration, calendar time), often in piecewise constant intensity (Poisson) models or semiparametric regression models such as the Cox model.

In the Poisson models the various time variables all enter the models as explanatory factors in a symmetrical way (and also symmetrical with respect to the other covariates of the model). However, in Cox models, one of the time variables must be chosen as the 'baseline' time variable while the others may be included as time-dependent covariates. The choice of baseline time variable may be governed by several considerations. First, the effect of the baseline time variable is given by the unspecified baseline hazard and, therefore, no regression coefficients are estimated for this variable. Thus, a time variable whose effect is of particular interest may not be the obvious choice as the baseline time variable. On the other hand, if a time variable is suspected to have an irregular effect which may not be easy to model parametrically via a time-dependent covariate, then this time variable may conveniently be chosen as the baseline time variable.[19]

An example concerning mortality of diabetics with or without nephropathy was studied by Andersen,[35] cf. the consequences for Danish life insurance premiums.[36]

The heterogeneity between individuals cannot always be accounted for by *observed* covariates, and at least some residual heterogeneity will have to be modelled as random variation. An important class of models here is the multiplicative *frailty models*, obtained by postulating random (usually transition-specific) *frailties* W^i_{hj}, and conditional transition intensities

$$\alpha^i_{hj}(t|W^i_{hj}) = \alpha_{0hj}(t)W^i_{hj}e^{\beta' Z_i}$$

Such models have been studied mainly for the survival data example, though some multi-state extensions are available.[4]

5.3 Hypothetical calculations in multi-state models

As mentioned earlier, there is often considerable interest in studying the consequences of the estimated transition intensities by calculating summary measures such as transition probabilities. When a full model has been estimated, this can be done not only for the model observed 'in this world'. Rather, the consequences of an assumed (or fitted) multi-state model may be usefully further illustrated by calculating transition probabilities in hypothetical models obtained by changing some of the parameters. An elaborate example of this was given by Keiding *et al.*[17] for the bone marrow transplantation context (see also the companion paper by Klein *et al.*[18]). Similar calculations have in fact been performed in the competing risks model ever since the first discussion by Bernoulli[37] of the effect on population mortality of removing smallpox through vaccination. The interpretational justification of such calculations was heatedly discussed, particularly in the 1970s.[7,10,38,39]

5.4 Partial model specification

We have so far assumed that the multi-state model was completely specified through statistical modelling of all transition intensities and a specific probability mechanism for the combination of these into transition probabilities. In a series of papers,[40–43] Pepe and her colleagues have developed estimates of certain functionals in multi-state models without assuming a full probability structure. One example is the *prevalence* of a transient condition indicated by state c defined as

$$\frac{P_{0c}(0,\ t)}{\sum_{j\in\mathcal{T}} P_{0j}(0,\ t)}$$

where $\mathcal{T}$ is the set of transient states and 0 is a fixed 'initial' state. The idea is to estimate numerator and denominator separately by simple linear combinations of Kaplan–Meier estimates. Easily applicable variance estimates are then available, which in one recent application[44] showed that the precision of the Pepe approach was close to the more elaborate (and restrictive) complete Markov model.

The product integral (Section 5.1) of the Nelson–Aalen estimator (3) has been studied by Datta and Satten[45,46] who showed that, also for non-Markovian processes, this combined with the initial distribution $\pi_h(0)$ provides consistent and asymptotically normal estimators for the state probabilities $\pi_h(t)$ (Section 3).

Andersen *et al.*[47] showed how regression models for transition probabilities $P_{hj}(s,\ t)$ or state probabilities $\pi_h(t)$ may be obtained directly in multi-state Markov models using jack-knife pseudo-observations (see also in the companion paper by Andersen *et al.*,[10] how this may be done for the special case of the competing risks model). In fact, their approach may be extended to state probabilities in non-Markovian models using the results of Glidden.[48]

Another example of partial model specification occurs when the model contains time-dependent covariates that are not purely endogenous. In fact, for time-fixed covariates we just conditioned (cf. Section 3) on their observed values without specifying their distribution $f(Z_i)$, but for time-dependent covariates such a conditioning is more tricky. Formally, the likelihood will contain factors for the stochastic development of $Z_i(t)$ given the history $\mathcal{F}_{t-} = \mathcal{X}_{t-} \vee \mathcal{Z}_{t-}$ and the likelihood (2) is not the full likelihood but only a partial likelihood for the parameters for the transition intensities $\alpha^i_{hj}(t)$. This means that inference for the transition *intensities* may be based on this partial likelihood whereas the transition *probabilities* will typically depend also on the parameters in the model for $Z_i(t)$.

Thus, if the model contains time-dependent covariates that are not purely endogenous then transition probabilities cannot be estimated using only a partial model specification. A joint model for the multi-state process $X_i(t)$ and the time-dependent covariates $Z_i(t)$ is needed. When $Z_i(t)$ only takes a finite number of values this joint model could, again, be a multi-state model where $Z_i(t)$ is now endogenous.[49,50] Examples of more general joint models were presented by Wulfsohn and Tsiatis[51] and Henderson *et al.*[52]

5.5 Example: PROVA

The main endpoints in the PROVA trial (Section 3) are bleeding events and deaths before bleeding, i.e., transitions out of state 0 on Figure 4. Figure 8 shows the Aalen–Johansen estimates for the probabilities of staying in state 0 for the four treatment groups in the trial. (This reduces, in fact, to the Kaplan–Meier estimator treating both bleedings and deaths before bleeding as failures, i.e., there are in total 96 events, cf. Table 1). It is seen that the estimated probabilities for patients who have received sclerotherapy (either alone or in combination with propranolol) tend to be lower than for other patients. Judged from a standard logrank test the differences between the four curves are close to significant ($P = 0.089$).

Analysing the bleeding intensity ($0 \rightarrow 1$) and the intensity of death without bleeding ($0 \rightarrow 2$) separately (Table 2) it is seen that it is the latter that is affected by treatment. Eliminating the three treatment indicators from a Cox regression model for $\alpha_{02}(t)$ also

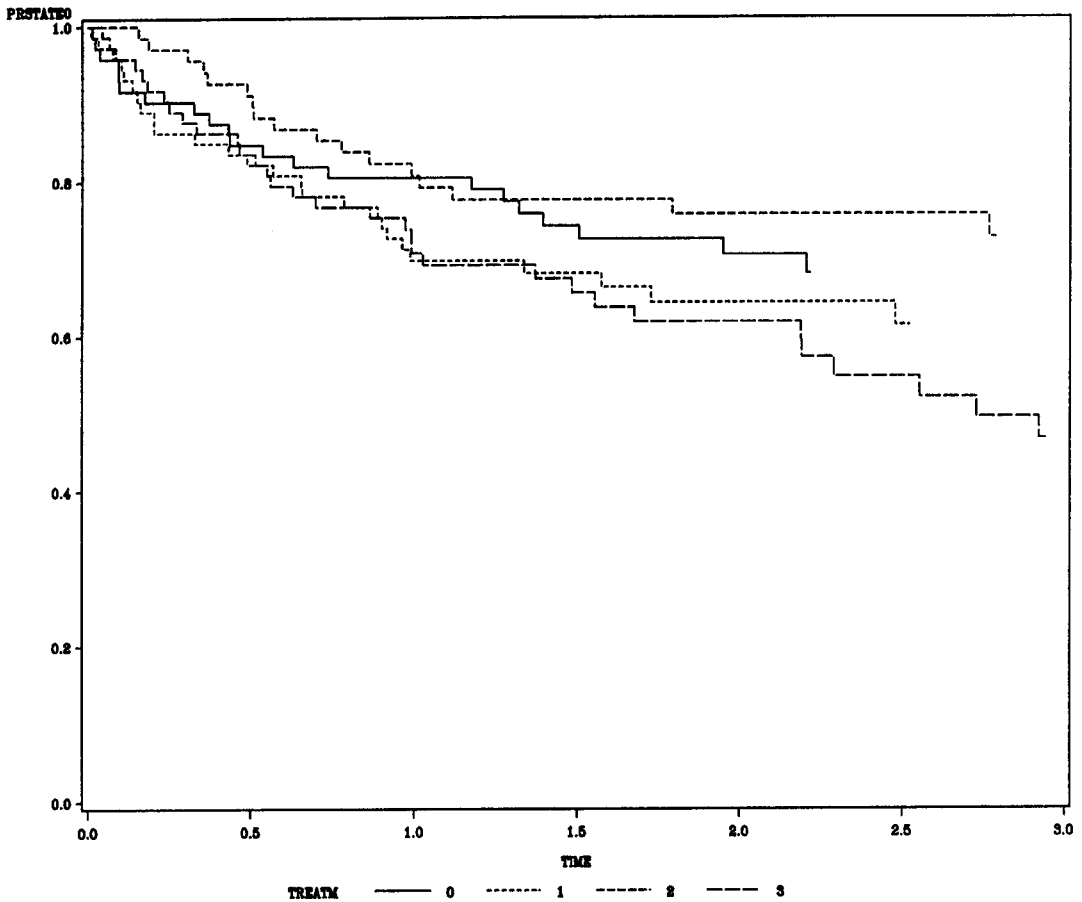

Figure 8 Estimates of staying in the bleeding-free state 0 in an illness–death model for the PROVA data, 0: control, 1: sclerotherapy, 2: propranolol, 3: both treatments.

Table 2 Estimates in Cox regression models for the PROVA data: bleeding intensity and intensity of death before bleeding

Covariate	$\alpha_{02}(t)$		$\alpha_{01}(t)$	
	$\hat{\beta}$	*SE*	$\hat{\beta}$	*SE*
Sclerotherapy	0.515	0.45	0.056	0.39
Propranolol	−0.319	0.57	−0.049	0.40
Both	0.908	0.42	−0.022	0.40
Sex(M = 1, F = 0)	0.928	0.42	0.154	0.32
Age(y)	0.0255	0.014	−0.0059	0.012

containing sex and age as explanatory variables is significant ($P = 0.029$) while a similar test for the bleeding intensity $\alpha_{01}(t)$ yields $P = 0.99$.

Turning to the $1 \rightarrow 2$ transition corresponding to death after bleeding there is a choice to make about which kind of model (Markovian or not) that fits the data. Table 3 (left column) shows estimates from an assumed Markovian model and it is seen that both sex and age seem to have significant effects. However, including time-dependent covariates for the sojourn time (*WAIT*) spent in state 1 (Table 3, right column) it is seen that, first of all, the Markov assumption is violated: the time-dependent covariates have a strong effect on the intensity ($P < 0.0001$) and, further, that the effects of sex and age are reduced when adjusting for *WAIT*. Therefore, a (semi-Markov) model with this time as the basic time variable was also fitted (Table 4). Here, neither the treatment variables, nor sex and age have significant effects and including time-dependent variables for time since randomization (*TIME*) has no effect either ($P = 0.79$, Table 4, right column). The explanation is that the mortality just after bleeding is very high and this is not captured well by the time-dependent variables in the model of Table 3. However, using the non-parametric baseline hazard in the Cox model to capture that effect gives a much more satisfactory model (Table 4).

5.6 Using standard software to analyse Cox-type multi-state regression models

Using the 'tricks' described in this section estimates for regression coefficients β_{hjm} and integrated baseline hazards $A_{hj0}(t) = \int_0^t \alpha_{hj0}(u)\mathrm{d}u$ in Cox-type models (4) may be obtained using standard packages like SAS, S-PLUS and STATA.

Table 3 Estimates in Cox regression models for the PROVA data: intensity of death after bleeding, time variable is time since randomization

Covariate	$\hat{\beta}$	*SE*	$\hat{\beta}$	*SE*
Sclerotherapy	−0.736	0.66	−0.775	0.66
Propranolol	0.190	0.63	0.174	0.65
Both	0.711	0.54	0.427	0.58
Sex(M = 1, F = 0)	1.321	0.50	1.258	0.53
Age(y)	0.0398	0.018	0.0316	0.019
5 days > *WAIT*			2.844	0.68
10 days > *WAIT* ≥ 5 days			2.216	0.77

Table 4 Estimates in Cox regression models for the PROVA data: intensity of death after bleeding, time variable is waiting time in state 1

Covariate	$\hat{\beta}$	*SE*	$\hat{\beta}$	*SE*
Sclerotherapy	−0.810	0.61	−0.810	0.62
Propranolol	−0.400	0.59	−0.398	0.59
Both	0.613	0.49	0.691	0.51
Sex(M = 1, F = 0)	0.697	0.51	0.707	0.51
Age(y)	0.0030	0.016	0.0017	0.018
1 year > *TIME*			0.207	0.83
2 years > *TIME* ≥ 1 year			0.0459	0.79

Recall from Section 5.1 that if all baseline hazards $\alpha_{hj0}(t)$ and all regression parameters β_{hjm} are distinct for the different transitions then the analysis may be performed one transition at a time using, for each transition $(h \to j)$ a data set with a record for each individual, i, including: T_{hi}^{in} = time of entry into state h for i, T_{hi}^{out} = time of exit from state h for i, D_{hji} = indicator for a $h \to j$ transition for i, $Z_{mi} = m_{hj}$ covariates for $i, m = 1, \ldots, m_{hj}$. Furthermore, if the desired model has non-proportional strata then S_{hji} = stratum for i, is also needed. Note that the covariates need not be the same for different (h, j), i.e., m_{hj} may depend on (h, j).

This has also been described by, for example, Therneau and Grambsch.[53] Here, we shall go a step further in order to analyse more parsimonious models where some baseline intensities are proportional or where some covariates have the same effect on several transition intensities. In such cases, a similar data set is required, but now one model is fitted to the large data set, which contains one record for each individual for each transition. The record for transition (h, j) should still contain T_{hi}^{in}, T_{hi}^{out} and D_{hji}. A stratum variable is also needed to separate baseline hazards, which are not assumed proportional. If two transition intensities are assumed proportional $\alpha_{hj}(t) = e^{\gamma}\alpha_{h'j'}(t)$ then they should have the same stratum value and a dummy 'covariate' should be added which is 1 for the data set corresponding to the $h \to j$ transition and 0 for the $h' \to j'$ transition *and all other transitions*. If a covariate Z_m has different effects for the $h \to j$ and the $h' - j'$ transitions and does not affect other transitions then the data set should contain two 'type-specific covariates' Z_m^1, Z_m^2 defined by

$$\begin{aligned} Z_m^1 &= Z_m, \; Z_m^2 = 0 && \text{for } h \to j \\ Z_m^1 &= 0, \quad Z_m^2 = Z_m && \text{for } h' \to j' \\ Z_m^1 &= 0, \quad Z_m^2 = 0 && \text{for all other transitions.} \end{aligned}$$

To exemplify, consider the following hypothetical three-state model for the PROVA data (Section 5.5, cf. Figure 4). Assume that:

(1) Death intensities (i.e., corresponding to transitions into state 2) are proportional, i.e., the baseline hazards are

$$\alpha_{12,0}(t) = e^{\gamma}\alpha_{02,0}(t)$$

(2) A binary covariate $Z_1 \sim \{1, 2\}$ gives rise to non-proportional bleeding ($0 \to 1$) intensities, i.e.,

$$\alpha_{01,0}(t) = \begin{cases} \alpha_{01,1}(t) & \text{if } Z_1 = 1 \\ \alpha_{01,2}(t) & \text{if } Z_1 = 2 \end{cases}$$

and Z_1 has no effect on the other transitions.

(3) A covariate Z_3 has the same effect (β_1) on the $0 \to 2$ and the $1 \to 2$ intensities but a different effect (β_2) on the $0 \to 1$ intensity.

(4) A covariate Z_3 has different effects (β_3, β_4) on the $0 \to 1$ and the $0 \to 2$ intensities and no effect on the $1 \to 2$ intensity.

This means that for individual i the model is:

$$\begin{aligned} \alpha^i_{02}(t) &= \alpha_{02,0}(t)\, e^{\beta_1 Z_{2i} + \beta_4 Z_{3i}} \\ \alpha^i_{01}(t) &= \alpha_{01,Z_{1i}}(t)\, e^{\beta_2 Z_{2i} + \beta_3 Z_{3i}} \\ \alpha^i_{12}(t) &= \alpha_{02,0}(t)\, e^{\gamma + \beta_1 \cdot Z_{2i}} \end{aligned} \tag{5}$$

and the three data records for this individual are given in Table 5. Here, everyone is assumed to be in state 0 at time 0 ($T^{in}_{0i} = 1$), though delayed entry is easily handled as mentioned in Section 4. Furthermore, T^{out}_{0i} is the time last seen in state 0, which is T_{01i} if a $0 \to 1$ transition is observed ($D_{01i} = 1$), T_{02i} if a $0 \to 2$ transition is observed ($D_{02i} = 1$), and the right-censoring time U_i if i is right-censored while being in state 0 ($D_{01i} = D_{02i} = 0$). If a $0 \to 1$ transition was observed then $T^{in}_{1i} = T_{01i}$ and T^{out}_{1i} is either the time of a $1 \to 2$ transition ($D_{12i} = 1$) or the right-censoring time $U_i (D_{12i} = 0)$. If no $0 \to 1$ transition is observed then both T^{in}_{1i} and T^{out}_{1i} are missing. Note that the stratum variable has three levels since there are three baseline hazards in the model; there is one dummy variable, the effect of which is the log hazard ratio γ between $\alpha_{12,0}(t)$ and $\alpha_{02,0}(t)$. This is, in fact, a purely endogeneous time-dependent covariate which is 1 when i is in state 1, and 0 otherwise. Finally, there are four type-specific covariates since the model has four unknown regression coefficients $\beta_1, \ldots, \beta_4$.

In this way estimates for the parameters for the transition intensities and their standard errors may be obtained. Such estimates may be combined to form transition probability estimates using the product integral as described by Andersen *et al.*[1] see also

Table 5 Data records for individual i for fitting the three-state Cox model (5)

Transition	T^{in}	T^{out}	D	Stratum	Dummy	Type-specific covariates			
						1	2	3	4
$0 \to 2$	0	T^{out}_{0i}	D_{02i}	02	0	Z_{2i}	0	0	Z_{3i}
$0 \to 1$	0	T^{out}_{0i}	D_{01i}	$\begin{cases} 01,1, & Z_{1i} = 1 \\ 01,2, & Z_{1i} = 2 \end{cases}$	0	0	Z_{2i}	Z_{3i}	0
$1 \to 2$	T^{in}_{1i}	T^{out}_{1i}	D_{12i}	02	1	Z_{2i}	0	0	0

the companion paper by Borgan.[54] However, except for the simple case of survival data no standard software exists for these computations (though, for the competing risks model, the companion paper by Andersen *et al.*[10] describes a SAS MACRO for this purpose).

6 Observational patterns

As emphasized in the introduction, event history data are rarely observed completely. Some patterns of incomplete observation are more easily handled than others, and this final section aims at introducing some of the more important classes. As mentioned above (cf. Section 4), independent delayed entry and right-censoring only modify the likelihood slightly and the statistical methods then all go through.

6.1 Interval censoring

An important class of incomplete observational patterns consists in the times of some (but often not all) transitions not being known exactly but only up to an interval, for example, between visits to a clinic or between censuses. (The *number* of transitions is usually assumed to be known precisely.) A classical approach in demography[55] is to approximate the 'exposure':

$$S_h^i = \int_{V_i}^{U_i \wedge \tau_i} Y_h^i(t)\,\mathrm{d}t$$

Another is to impute values in the observation interval for the time at risk. There are a number of systematic studies of nonparametric maximum likelihood estimation under interval censoring of the healthy → diseased transition in the unidirectional illness–death model.[56–59] Kay[60] and Andersen *et al.*[61] exemplified interval-censored observation in the reversible illness–death model, using piecewise constant intensity models. A particular example is *panel data*, see, for example Gentleman *et al.*[62] and the references therein. Interval-censoring in multi-state models is discussed in the companion paper by Commenges.[63]

6.2 Conditioning in multi-state models

Many observational patterns in event history analysis may be described by conditional distributions in those simple models, which often describe 'direct' observations that are practically unobtainable.[64–66] A prime example is the left truncation just described, another is *right truncation*, with widespread use in studies of AIDS patients whose development is often observed conditional on having contracted the disease before the study entry. A more elaborate application of such retrospective observational plans obtained by conditioning in an underlying 'prospective' Markov process model was documented by Aalen *et al.*[15] for the four-state interaction of life history events example described above (Figure 6). These authors relied heavily on the concise but important general framework of Hoem.[67] We shall here briefly outline how this methodology works for a simple example of retrospective incidence estimation,

obtained from the Markov illness–death model without recovery illustrated in Figure 4. Assume that we study a random sample of individuals alive at same fixed age u; for those who had by then contracted the disease the age at which this happened is recorded. The observed multi-state model has state space $\mathcal{K} = \{0, 1\}$ and transition probabilities

$$Q_{hj}(s, t) = \text{Prob}\{X(t) = j | X(s) = h, X(u) \neq 2\}$$

for $h, j \in \{0, 1\}$ and $0 < s < t < u$. We get

$$Q_{hj}(s, t) = P_{hj}(s, t) \frac{P_{j\mathcal{K}}(t, u)}{P_{h\mathcal{K}}(t, u)}$$

where $P_{hj}(s, t)$ are the transition probabilities in the original illness–death process and $P_{h\mathcal{K}} = P_{h0} + P_{h1}$. Hoem[67] used the term *purged* for the conditional Markov process on $\mathcal{K}$ with transition probabilities $Q_{hj}(s, t)$. The transition intensity of the purged process is

$$\lambda_{01}(t) = \alpha_{01}(t) \frac{P_{11}(t, u)}{P_{1\mathcal{K}}(t, u)}$$

and it may be proved that if the mortality of the diseased is never smaller than that of the healthy, i.e., $\alpha_{02}(t) \leq \alpha_{12}(t)$ for all $t \leq u$, then $\lambda_{01}(t) \leq \alpha_{01}(t)$, with equality if and only if $\alpha_{02}(t) \equiv \alpha_{12}(t)$. This documents the intuitively obvious result, that the retrospective study will underestimate the disease incidence because of *survivor selection*. Andersen and Green[68] used such methodology to study robustness of diabetes incidence estimates in a situation where diabetics were only observed conditionally on not emigrating before a certain age.

6.3 The prevalent cohort study

An important sampling frame for the illness–death model without recovery is the *prevalent cohort study* where a cross-sectional sample of diseased is taken at a fixed calendar time. Keiding,[69] cf. Lund,[70] discussed the conditions for correct inference on mortality $\alpha_{12}(t)$ based on follow-up of the diseased, studied under left truncation, and compared to inference based on the *length-biased* durations, which include the retrospective time from disease onset, as well as to the *forward recurrence time* from sampling to death, assuming stationarity.

A comprehensive substantive discussion of biases in prevalent cohorts was given by Brookmeyer.[71] Retrospective estimation of incidence based on the disease onset information of the survivors and independent lethality information was exemplified by Keiding *et al.*,[72] cf. Ogata *et al.*[73]

6.4 Some other partial information designs

Sometimes interval censoring is extreme: in a cross-sectional study it is for all individuals only known whether or not an event has happened at age of sampling. Such *current-status data* were discussed in detail by Diamond and

McDonald,[74] Keiding[75] and Keiding *et al.*[76]; and there is an elaborate recent mathematical-statistical development in this area, see for example, Groeneboom and Wellner[77] and Lin *et al.*[78] For *time to pregnancy* data, Keiding *et al.*[29] proposed using the *current duration* elapsed so far; under suitable stationarity conditions this distribution can be considered a backward recurrence time.

Acknowledgements

This research was partly supported by Grant RO1-CA54706-09 from the National Cancer Institute and by a grant from the Danish National Research Foundation.

References

1 Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical models based on counting processes.* New York: Springer, 1993.
2 Blossfeld H, Rohwer G. *Techniques of event history modeling.* New Jersey: Lawrence Erlbaum, 1995.
3 Courgeau D, Lelièvre E. *Event history analysis in demography.* Oxford: Clarendon, 1992.
4 Hougaard P. *Analysis of multivariate survival data.* New York: Springer, 2000.
5 Hougaard P. Multi-state models: A review. *Lifetime Data Analysis* 1999; **5**: 239–64.
6 Commenges D. Multi-state models in epidemiology. *Lifetime Data Analysis* 1999; **5**: 315–27.
7 Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data.* New York: Wiley, 1980.
8 Arjas E, Haara P. A marked point process approach to censored failure data with complicated covariates. *Scandinavian Journal of Statistics* 1984; **11**: 193–209.
9 Hoem JM. The statistical theory of demographic rates. A review of current developments (with discussion). *Scandinavian Journal of Statistics* 1976; **3**: 169–85.
10 Andersen PK, Abildstrom S, Rosthøj S. Competing risks as a multi-state model. *Statistical Methods in Medical Research* 2002; **11**: 203–15.
11 Fix E, Neyman J. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* 1951; **23**: 205–41.
12 Sverdrup E. Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. *Skandinavisk Aktuarietidsskrift* 1965; **48**: 184–211.
13 PROVA Study Group. Prophylaxis of first hemorrhage from esophageal varices by sclerotherapy, propranolol or both in cirrhotic patients: a randomized multicenter trial. *Hepatology* 1991; **14**: 1016–24.
14 Cook RJ, Lawless JF. Analysis of repeated events. *Statistical Methods in Medical Research* 2002; **11**: 141–66.
15 Aalen OO, Borgan Ø, Keiding N, Thormann J. Interaction between life history events: nonparametric analysis of prospective and retrospective data in the presence of censoring. *Scandinavian Journal of Statistics* 1980; **7**: 161–71.
16 Klein JP, Keiding N, Copelan EA. Plotting summary predictions in multistate survival models: Probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine* 1993; **12**: 2315–32.
17 Keiding N, Klein JP, Horowitz MM. Multistate models and outcome prediction in bone marrow transplantation. *Statistics in Medicine.* 2001; **20**: 1871–85.
18 Klein, JP, Shu Y. Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research* 2002; **11**: 117–139.
19 Clayton D, Hills M. *Statistical models in epidemiology.* Oxford: Oxford University Press, 1993.
20 Lindsay JC, Ryan LM. A three-state multiplicative model for rodent tumorigenicity experiments. *Applied Statistics* 1993; **42**: 283–300.
21 Chiang CL. *Introduction to stochastic processes in biostatistics.* New York: John Wiley & Sons, Inc, 1968.
22 Aalen OO. Statistical inference for a family of counting processes. PhD thesis, University of California, Berkeley, 1975.

23 Aalen OO. Nonparametric inference for a family of counting processes. *Annals of Statistics* 1978; **6**: 701–26.

24 Aalen OO, Johansen S. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* 1978; **5**: 141–50.

25 Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**: 457–81.

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

27 Cox DR. The statistical analysis of dependencies in point processes. In: Lewis PAW, ed. *Stochastic point processes.* New York: John Wiley and Sons, 1972.

28 Cox DR. Partial likelihood. *Biometrika* 1975; **62**: 269–76.

29 Keiding N, Hartvig H, Tvede M, Juul S. Estimating time to pregnancy from current durations in a cross-sectional sample. Research Report 99/7, Department of Biostatistics, University of Copenhagen, 1999.

30 Pierce DA, Preston DL. Joint analysis of site-specific cancer risks for the atomic bomb survivors. *Radiation Research* 1993; **134**: 134–42.

31 Wohlfahrt J, Andersen PK, Melbye M. Multivariate competing risks. *Statistics in Medicine* 1999; **18**: 1023–30.

32 Aalen OO. A model for non-parametric regression analysis of counting processes. *Springer Lecture Notes Statistics* 1980; **2**, 1–25. In: Klonecki W, Kozek A, Rosiński J, eds. *Mathematical statistics and probability theory*.

33 Aalen OO. A linear regression model for the analysis of life times. *Statistics in Medicine* 1989; **8**: 907–25.

34 Aalen OO, Borgan Ø, Fekjær H. Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* 2001; **57**: 993–1001.

35 Andersen PK. Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine* 1988; **7**: 661–70.

36 Ramlau-Hansen H, Jespersen NCB, Andersen PK, Borch-Johnsen K, Deckert T. Life insurance for insulin-dependent diabetics. *Scandinavian Actuarial Journal* 1987; 19–36.

37 Bernoulli D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour le prévenir. *Historie avec le Mémorires, Académie Royal des Sciences* 1760, Paris, 1–45.

38 Gail M. A review and critique of some models used in competing risk analysis. *Biometrics* 1975; **31**: 209–22.

39 Prentice RL, Kalbfleisch JD, Peterson AV *et al.* The analysis of failure time data in the presence of competing risks. *Biometrics* 1978; **34**: 541–54.

40 Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association* 1991; **86**, 770–78.

41 Pepe MS, Longton G, Thornquist M. A qualifier *Q* for the survival function to describe the prevalence of a transient condition. *Statistics in Medicine* 1991; **10**: 413–21.

42 Pepe MS, Mori M. Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* 1993; **12**: 737–51.

43 Couper D, Pepe MS. Modelling prevalence of a chronic condition: chronic graft-versus-host disease after bone marrow transplantation. *Statistics in Medicine* 1997; **16**: 1551–71.

44 Klein JP, Keiding N, Shu Y, Szydlo RM, Goldman JM. Summary curves for patients transplanted for chronic myeloid leukemia salvaged by a donor lymphocyte infusion: the current leukemia free survival curve. *British Journal of Haematology* 2000; **109**: 148–52.

45 Datta S, Satten GA. Validity of the Aalen Johansen estimators of stage occupation probabilities and Nelson Aalen integrated transition hazards for Non-Markov models. *Statistics and Probability Letters* 2001, **55**: 403–411.

46 Satten GA, Datta S. Marginal estimation for multi-state models: waiting time distributions and competing risks analyses. *Statistics in Medicine* 2002; **21**: 3–19.

47 Glidden DV. Robust inference for event probabilities with non-Markov event data. *Biometrics* (in press).

48 Andersen PK. Time-dependent covariates and Markov processes. In: Moolgavkar SH, Prentice RL, eds. *Modern statistical methods in chronic disease epidemiology*, New York: John Wiley and Sons, 1986.

49 Andersen PK, Hansen LS, Keiding N. Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scandinavian Journal of Statistics* 1991; **18**: 153–67.

50 Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–39.

51 Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**: 465–80.

52 Therneau TM, Grambsch PM. *Modeling survival data. Extending the cox model.* New York: Springer, 2000.

53 Borgan Ø. Estimation of covariate-dependent Markov transition probabilities from nested case-control data. *Statistical Methods in Medical Research* 2002; **11**: 183–202.

54 Andersen PK, Klein JP, Rosthøj S. From summary statistics to generalized linear models for pseudo-observation, with applications to multi-state models. *Submitted to Biometrika.*

55 Hoem JM. Fertility rates and reproduction rates in a probabilistic setting. *Biométrie-Praximétrie* 1969; **10**: 38–66. Correction, 1970; **11**: 20.

56 Frydman H. A non-parametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *Journal of the Royal Statistical Society, Series B* 1992; **54**: 853–66.

57 Frydman H. Semiparametric estimation in a three-state duration dependent Markov model from interval-censored observations with application to AIDS data. *Biometrics* 1995; **51**: 502–11.

58 Joly P, Commenges D. A penalized approach for a progressive three-state model with censored and truncated data: Application to AIDS. *Biometrics* 1999; **55**: 887–90.

59 Gaüzère F. Approche non-paramétrique pour un modèle 3 états avec censures par intervalles—application à la dépendance. PhD thesis, Université Victor Segalen Bordeaux 2, 2000.

60 Kay R. A Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics* 1986; **42**: 855–65.

61 Andersen PK, Hansen LS, Keiding N. Assessing the influence of reversible disease indicators on survival. *Statistics in Medicine* 1991; **10**: 1061–67.

62 Gentleman RC, Lawless JF, Lindsay JC, Yan P. Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine* 1994; **13**: 805–21.

63 Commenges D. Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research* 2002; **11**: 167–82.

64 Kalbfleisch JD, Lawless JF. Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association* 1989; **84**: 360–72.

65 Keiding N, Gill RD. Random truncation models and Markov processes. *Annals of Statistics* 1990; **18**: 582–602.

66 Gross ST, Huber-Carol C. Regression models for right truncated left censored survival data. *Scandinavian Journal of Statistics* 1992; **19**: 193–213.

67 Hoem JM. Purged and partial Markov chains. *Skandinavisk Aktvarietidsskrift* 1969; **52**: 147–55.

68 Andersen PK, Green A. Evaluation of estimation bias in an illness–death–emigration model. *Scandinavian Journal of Statistics* 1985; **12**: 63–68.

69 Keiding N. Independent delayed entry. In: Klein JP, Goel PK, eds. *Survival analysis: state of the art.* Dordrecht: Kluwer, 1992.

70 Lund J. Sampling bias in population studies—how to use the Lexis diagram. *Scandinavian Journal of Statistics* 2000; **27**: 589–604.

71 Brookmeyer R. Biased sampling of cohorts in epidemiology. In: Armitage P, Colton T, eds. *Encyclopedia of biostatistics.* New York: John Wiley and Sons, 1998.

72 Keiding N, Holst C, Green A. Retrospective estimation of diabetes incidence from information in a current prevalent population and historical mortality. *American Journal of Epidemiology* 1989; **130**: 588–600.

73 Ogata Y, Katsura K, Keiding N, Holst C, Green A. Empirical Bayes age-period-cohort analysis of retrospective incidence data. *Scandinavian Journal of Statistics* 2000; **27**: 415–432.

74 Diamond ID, McDonald JW. Analysis of current-status data. In: Trussel J, Hankinson R, Tilton J, eds. *Demographic applications of*

event history analysis. Oxford: Clarendon Press, 1992.
75 Keiding N. Age-specific incidence and prevalence: a statistical perspective (with discussion). *Journal of the Royal Statistical Society, Series A* 1991; **154**: 371–412.
76 Keiding N, Begtrup K, Scheike TH, Hasibeder G. Estimation from current-status data in continuous time. *Lifetime Data Analysis* 1996; **2**: 119–29.
77 Groeneboom P, Wellner JA. *Information bounds and nonparametric maximum likelihood estimation.* Basel: Birkhäuser, 1992.
78 Lin DY, Oakes D, Ying Z. Additive hazards regression with current status data. *Biometrika* 1998; **85**: 289–98.