Noname manuscript No.
(will be inserted by the editor)

Inference for outcome probabilities in multi-state models

Per Kragh Andersen · Maja Pohar Perme

Received: date / Accepted: date

Abstract In bone marrow transplantation studies, patients are followed over time and a number of events may be observed. These include both ultimate events like death and relapse and transient events like graft versus host disease and graft recovery. Such studies, therefore, lend themselves for using an analytic approach based on multi-state models. We will give a review of such methods with emphasis on regression models for both transition intensities and transition- and state occupation probabilities. Both semi-parametric models, like the Cox regression model, and parametric models based on piecewise constant intensities will be discussed.

Keywords multi-state models · bone marrow transplant · survival analysis · regression models

1 Introduction

For diseases like acute or chronic leukemia, and also for a range of other diseases, bone marrow transplantation (BMT) is an effective and frequently used type of treatment, e.g. Klein and Shu (2002). The ultimate outcome of interest in such studies is the death of the patient, however, relapse of the disease is a sign of treatment failure as well, and it is also frequently considered an ultimate outcome. In addition to these "absorbing" states, death and relapse,

P.K. Andersen
Department of Biostatistics, University of Copenhagen
O. Farimagsgade 5, PB 2099, DK 1014 Copenhagen K, Denmark
Tel.: +45-353-27908
Fax: +45-353-27907
E-mail: P.K.Andersen@biostat.ku.dk

M. Pohar Perme
Department of Biomedical Informatics, University of Ljubljana,
Vrazov trg 2, SI-1000 Ljubljana, Slovenia.

a number of intermediate ("transient") states are recognized in BMT studies. These include acute and chronic graft versus host disease (GvHD) and graft recovery.

For these reasons, multi-state models (MSM) are extremely useful in the analysis of follow-up data from BMT studies. This paper provides a review of the area. Previous summary papers on MSMs include Hougaard (1999), Commenges (1999), Andersen and Keiding (2002), Putter et al. (2007) and Meira-Machado et al. (2009).

A *multi-state process* is a (continuous-time) stochastic process $(X(t), t \in \mathcal{T})$ with a finite *state space* $\mathcal{S} = \{0, 1, \ldots, p\}$ and with right-continuous sample paths: $X(t+) = X(t)$. Here, $\mathcal{T} = [0, \tau]$ or $[0, \tau)$ with $\tau \leq +\infty$. 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

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

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

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

which we shall assume exist. Some transition intensities may be 0 for all t. The *state occupation probabilities* are $\pi_h(t) = \text{Prob}(X(t) = h), h \in \mathcal{S}$ and, in particular, the *initial distribution* is $\pi_h(0) = \text{Prob}(X(0) = h), h \in \mathcal{S}$. We may then write

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

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. We shall illustrate this in connection with the example below.

A state $h \in \mathcal{S}$ is *absorbing* if for all $t \in \mathcal{T}, j \in \mathcal{S}, j \neq h, \alpha_{hj}(t) = 0$, that is, no arrows in the diagram begin in h; otherwise h is *transient*. 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 time 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 $Z(t)$ may give rise to an extended history of the form $\mathcal{F}_t = \mathcal{X}_t \vee \mathcal{Z}_t$ where $\mathcal{Z}_t$ is the history generated by the covariates in $[0, t]$.

Thus, a number of parameters are of interest in MSMs: transition intensities, transition probabilities, state occupation probabilities, and distributions of time spent in each state. Among these, probabilities have a more direct interpretation and this paper will focus on inference for probabilities. However, since inference for intensities is often more simple, and since some models for

Table 1 Description of data from 2009 patients who underwent bone marrow transplantation.

Age (mean, SD)	31.9	15.4
Female sex	896	44.6%
Disease type AML	1406	70.0%
Graft type BM only	1153	57.4%
Karnofsky score (mean, SD)	91.3	9.1 (25 missing values)
Relapse	259	12.9%
Death	737	36.7%
Death and relapse	232	89.6% of patients with relapse
AGvHD	590	29.4%
CGvHD	630	31.4%
A or C GvHD	989	49.2%
A or C GvHD and relapse	104	10.9% of patients with GvHD
A or C GvHD and death	400	40.4% of patients with GvHD

probabilities are defined by plugging-in models for intensities we will first give a brief summary of intensity models.

Throughout, the methods to be discussed will be illustrated using a transplant outcome data set from The Center for International Blood and Marrow Transplant Research (CIBMTR). The CIBMTR is comprised of clinical and basic scientists who confidentially share data on their blood and bone marrow transplant patients with CIBMTR Data Collection Center located at the Medical College of Wisconsin. The CIBMTR is a repository of information about results of transplants at more than 450 transplant centers worldwide. The example data set consists of patients who received HLA-identical sibling transplant from 1995 to 2004 for acute myelogenous leukemia (AML) or acute lymphoblastic leukemia (ALL) and were transplanted in first complete remission. All patients received bone marrow transplantation or peripheral blood stem cell transplantation. The infants aged less than 2 years old and all patients who received umbilical cord blood transplants were excluded as risk factors are likely to vary in this group. Table 1 presents a description of the example data set.

2 Models and inference for transition intensities

An attractive feature of multi-state models based on transition intensities is that all hazard-based models known from survival analysis apply, see e.g. Andersen et al. (1993). This includes both estimation, testing and model checking. For ease of notation we first fix two states, $h, j \in \mathcal{S}$, and denote the $h \to j$ transition intensity $\alpha(t)$.

2.1 Models for homogeneous populations

We first study the case of no covariates where there are, basically, the two options of either applying non-parametric or parametric models. In the former,

$\alpha(t)$ is left completely unspecified while, in the latter, the simplest possible model is the constant hazard model, $\alpha(t) = \alpha$. Since the constant hazard assumption is often too simple to apply in practice, a frequently used extension of this is the piecewise constant hazard model where, for some time-intervals given by cut-points $0 = \tau_0 < \tau_1 < \ldots < \tau_K = \tau$, it is assumed that $\alpha(t) = \alpha_\ell$ for $\tau_{\ell-1} \leq t < \tau_\ell, \ell = 1, \ldots, K$. Another extension of the simple constant hazard model is the Weibull model with $\alpha(t) = \alpha t^\gamma, \gamma \geq 0$ which for $\gamma = 0$ reduces to the constant hazard model.

The models discussed so far are all *Markovian* as the transition intensity at time t does not depend on other aspects of the past history, $\mathcal{X}_{t-}$ than the state (h) occupied at $t-$. This is an important class of models where, as we shall see in later sections, transition *probabilities* may be derived from the intensities. Another important class of models is *semi-Markov* models where the $h \to j$ transition intensity at time t also depends on the *duration* or *sojourn time* in state h, that is $t - T$ where $T(\leq t)$ is the time of entry into state h. The constant hazard model is, obviously, always Markov. A model where there is only duration dependence (and no direct dependence on the baseline time variable, "calendar time", t) is some times called *homogeneous semi-Markov* to distinguish it from a general semi-Markov model where $\alpha(\cdot)$ may depend on both calendar time, t and duration, $t - T$. We shall return to general semi-Markov models below when regression models are to be introduced.

2.2 Regression models

Most regression models involve a linear predictor, that is a linear function of the covariates for individual i, $i = 1, \ldots, n$, with some unknown regression coefficients, β_m, (or regression functions, $\beta_m(t)$):

$$\mathrm{LP}_i(t) = \sum_{m=1}^{k} \beta_m(t) Z_{mi}(t)$$

where covariates are allowed to be time-dependent. We shall restrict attention to regression models with a linear predictor. This means that, when covariates are to be included into models for transition intensities, a choice of link function has to be made, i.e. one needs to specify how $\mathrm{LP}_i(t)$ relates to $\alpha(t \mid Z_i)$. Many hazard models are multiplicative, i.e. $\log(\alpha(t \mid Z_i))$ is linear in the covariates, but also additive hazard models (where the link is the identity function) are frequently studied, see e.g. Martinussen and Scheike (2006).

In what follows we will only discuss multiplicative models which means that, when introducing covariates into the non-parametric model above, the semi-parametric Cox regression model

$$\alpha(t \mid Z_i) = \alpha_0(t) \exp(\mathrm{LP}_i(t)) \tag{1}$$

is obtained. Here, the baseline hazard $\alpha_0(t)$ is left completely unspecified while the regression coefficients, β_m are usually assumed constant leading to the

basic proportional hazards assumption of this model. The covariates, however, may be time-dependent.

In a similar way, covariates may be introduced into parametric hazard models. This leads to fully parametric models for $\alpha(t \mid Z_i)$ of the same multiplicative form as the Cox model (1) and with the baseline hazard, $\alpha_0(t)$ being, e.g. constant or piecewise constant.

As for the models for homogeneous populations, the regression models lead to Markov processes if the hazard at ("calendar") time t does not depend on other aspects of $\mathcal{F}_{t-}$ than the state (h) occupied at $t-$, that is, if the baseline hazard is a function of t and if no function of the past is included as a time-dependent covariate. Similarly, the process is homogeneous semi-Markov if the hazard at time t only depends on the duration, $t-T$, of the stay in the current state, h.

However, the allowance for time-dependent covariates provides a means for studying general semi-Markov processes where the hazard at time t depends on both time variables, t and $t-T$. In the framework of the semi-parametric Cox regression model this is done by choosing one of the time variables as the "baseline" time variable, i.e. $\alpha_0(\cdot)$ is a function of that time variable, while functions of the other time variable may be included as time-dependent covariates, e.g.

$$\alpha(t \mid Z_i) = \alpha_0(t) \exp(\beta_0 f(t - T_i) + \mathrm{LP}_i(t)),$$

with $f(\cdot)$ being some pre-specified function of the current duration like the identity $f(d) = d$ or the indicator $f(d) = I(d \leq d_0)$ corresponding to some duration threshold, d_0. Such a model is frequently used to test the Markov null hypothesis: $\beta_0 = 0$.

2.3 Inference for transition intensities

We will assume that independent multi-state processes

$$(X_i(t), 0 \leq t \leq C_i; i = 1, \ldots, n)$$

are observed in continuous time, that is, times of transition are observed exactly, except for the fact that independent right-censoring at $C_i (\leq \tau)$ is allowed for $X_i(\cdot)$. In the next section, interval-censoring will be touched upon. Left-truncation may quite easily be incorporated, as well, but for sake of simplicity we have chosen not to do so since that would require specific assumptions concerning the information available at the time of left-truncation. The data for individual i can then be represented as a multivariate counting process

$$N_{hji}(t), h, j \in \mathcal{S}, h \neq j, t \leq C_i$$

counting the number of direct $h \to j$ transitions observed for subject i in $[0, t]$ (where some h, j combinations may not be possible). The model is then

specified by the transition intensities $\alpha_{hji}(t \mid Z_i(t))$ leading to the intensity process

$$\lambda_{hji}(t) = Y_{hi}(t)\alpha_{hji}(t \mid Z_i(t))$$

for $N_{hji}(t)$ with respect to the filtration $(\mathcal{F}_t)$ and the relevant probability measure. That is,

$$\mathrm{E}(N_{hji}(t) \mid \mathcal{F}_{t-}) = \int_0^t \lambda_{hji}(u)\mathrm{d}u. \tag{2}$$

Here,

$$Y_{hi}(t) = I(X_i(t-) = h)$$

is the indicator of $X_i(\cdot)$ being in state h at time $t-$. For a filtration of the form $\mathcal{F}_t = \mathcal{X}_t \vee \mathcal{Z}_0$, the likelihood for the model parameters, say $\boldsymbol{\theta}$, is obtained via Jacod's formula (e.g. Andersen et al. 1993, Chapter II):

$$L(\boldsymbol{\theta}) = \prod_i \prod_{h,j} \left(\prod_t \lambda_{hji}(t)^{\Delta N_{hji}(t)} \right) \exp(-\int_0^{C_i} \lambda_{hji}(u)\mathrm{d}u). \tag{3}$$

Strictly speaking, (3) is only a partial likelihood since potential randomness in the right-censoring times, C_i, is not accounted for. The filtration will have the form $\mathcal{X}_t \vee \mathcal{Z}_0$ when, either, only time-fixed covariates are included, or when time-dependent covariates only depend on the past history of the multi-state process because in these cases the covariates no not impose extra randomness for $t > 0$. In the case of an extended filtration $\mathcal{F}_t = \mathcal{X}_t \vee \mathcal{Z}_t$, that is, when more general time-dependent covariates are allowed, (3) is only a partial likelihood, see e.g. Andersen et al. (1993, Chapters II-III) for further discussion.

At any rate, (3) may be used as the basis for inference on $\boldsymbol{\theta}$. For the purely non-parametric model: $\alpha_{hji}(t) = \alpha_{hj}(t)$, completely unspecified, this leads (with proper definition of non-parametric maximum likelihood estimation) to the *Nelson-Aalen* estimator

$$\widehat{A}_{hj}(t) = \int_0^t \frac{\mathrm{d}N_{hj}(u)}{Y_h(u)} \tag{4}$$

for the cumulative transition intensity

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

In (4), $N_{hj}(t)$ and $Y_h(t)$ are the aggregated processes

$$N_{hj} = \sum_i N_{hji}, \quad Y_h = \sum_i Y_{hi}.$$

For the piecewise constant hazard model

$$\alpha_{hji}(t) = \alpha_{hj\ell} \text{ for } \tau_{hj,\ell-1} \leq t < \tau_{hj,\ell}, \ell = 1, \ldots, K_{hj},$$

(3) leads to the following "occurrence/exposure rate" estimators

$$\widehat{\alpha}_{hj\ell} = \frac{D_{hj\ell}}{T_{h\ell}}$$

with $D_{hj\ell} = N_{hj}(\tau_{hj,\ell}-) - N_{hj}(\tau_{hj,\ell-1})$ being the total number of $h \to j$ transitions in interval ℓ and

$$T_{h\ell} = \int_{\tau_{hj,\ell-1}}^{\tau_{hj,\ell}} Y_h(u)\mathrm{d}u$$

the total time at risk for such transitions in that interval.

Variance estimates are available from the second derivative of $-\log L(\boldsymbol{\theta})$, though robust standard errors may also be applied (e.g. Lin et al. 2000).

For the Cox regression model

$$\alpha_{hji}(t \mid Z_i(t)) = \alpha_{hj0}(t)\exp(\mathrm{LP}_{hji}(t))$$

a number of options is available for specification of the linear predictor. The most obvious choice is to let both the covariates and the regression coefficients vary between transition types, i.e. to let

$$\mathrm{LP}_{hji}(t) = \sum_{m=1}^{k_{hj}} \beta_{hjm} Z_{hjmi}(t).$$

However (see, e.g. Andersen et al. 1993, Chapter VII, for further discussion), working with a single ("long") vector, $(\beta_1, \ldots, \beta_k)$, of regression coefficients and defining, appropriately, type specific covariates $Z_{hjmi}(t), m = 1, \ldots, k$, provides added flexibility in that models where some coefficients are shared between several transition types may be studied. This specification of the linear predictor

$$\mathrm{LP}_{hji}(t) = \sum_{m=1}^{k} \beta_m Z_{hjmi}(t)$$

leads to the following version of the Cox partial likelihood

$$\mathrm{CL}(\beta) = \prod_t \prod_{hji} \left(\frac{\exp(\mathrm{LP}_{hji}(t))}{\sum_r Y_{hr}(t)\exp(\mathrm{LP}_{hjr}(t))} \right)^{\Delta N_{hji}(t)} \tag{5}$$

as a profile likelihood from (3) obtained by partial maximization over $A_{hj0}(\cdot)$. When no regression coefficients are common to different transition intensities, (5) is a product over transitions types, h, j, with separate parameters for the different types, i.e. maximization may be performed for one type of transition at a time. In general, the whole Cox partial likelihood (5) must be maximized in one analysis (utilizing the concept of *stratified* Cox models, e.g., Andersen et al., 1993, Chapter VII; Andersen and Keiding, 2002). Furthermore, the cumulative baseline hazard is estimated by the Breslow estimator

$$\widehat{A}_{hj0}(t) = \int_0^t \frac{\mathrm{d}N_{hj}(u)}{\sum_r Y_{hr}(u)\exp(\widehat{\mathrm{LP}}_{hjr}(u))}.$$

For the parametric regression model with piecewise constant transition intensities, (3) leads to so-called *Poisson* regression. When covariates are all discrete, that is, when the covariate vector takes values, v, in a finite set V, the data may be summarized as the "tables":

$$(D_{hj\ell,v}, T_{h\ell,v}, v \in V, \ell = 1, \ldots, K_{hj}).$$

Here, $D_{hj\ell,v}$ are the $h \to j$ transition counts and $T_{h\ell,v}$ the time at risk in h in each cell obtained by a cross-classification of v and time-interval, ℓ and these tables are *sufficient*. Furthermore, the likelihood is proportional to that obtained treating $D_{hj\ell,v}$ as Poisson with a mean

$$\alpha_{hj\ell,v} T_{h\ell,v}$$

proportional to $T_{h\ell,v}$. Here, the $h \to j$ transition intensity $\alpha_{hj\ell,v}$ is a piecewise constant function of time and, in the multiplicative Poisson model, $\log(\alpha_{hj\ell,v})$ is linear in the covariates.

For large samples, this sufficiency reduction of the follow-up data may lead to considerable simplification of the inference without seriously losing the flexibility of the non-parametric baseline hazard in the Cox model (though, of course, a choice of time-intervals needs to be made). Furthermore, Poisson regression has the advantages that non-proportional hazards is simply an interaction between time and covariates and, further, that non-homogeneous semi-Markov models are easily handled by splitting follow-up time according to both ("calendar") time t and duration, $t - T$. This means that, when analyzing such models, a choice of "baseline time-variable" (calendar time or duration) is not needed since both appear in the log-hazard model on equal footings. Finally, the parameters for effects of such time variables will be part of the standard regression output from any computer package in contrast to the Breslow estimates from the Cox models which will usually appear in the output, in either tabular or graphical form, only when specifying the relevant options.

2.4 Inference for marginal rate functions

The basic counting processes, $N_{hji}(t)$, have intensity processes, $\lambda_{hji}(t) = Y_{hi}(t)\alpha_{hji}(t)$ directly depending on the transition intensities, $\alpha_{hji}(t)$. By (2), these are conditional expectations

$$\mathrm{E}(N_{hji}(t) \mid \mathcal{F}_{t-}) = \int_0^t \lambda_{hji}(u)\mathrm{d}u$$

given the entire history $\mathcal{F}_{t-}$. Some times, one may be interested in studying only certain marginal properties of the multi-state process, e.g. the marginal means

$$\mathrm{E}(N_{hji}(t) \mid Z_i) = \int_0^t \widetilde{\lambda}_{hji}(u)\mathrm{d}u,$$

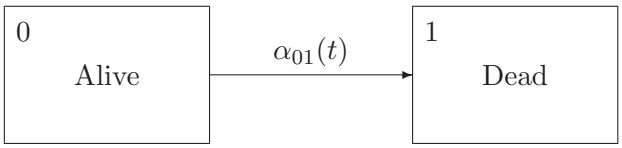


Fig. 1 The two-state model for survival data.

only conditioning of time-fixed covariates, Z_i or, more generally,

$$\mathrm{E}(N_{hji}(t) - N_{hji}(s) \mid X_i(s) = h, Z_i) = \int_s^t \widetilde{\lambda}_{hji}(u)\mathrm{d}u,$$

only conditioning on being in state h at $s-$ (and on time-fixed covariates, Z_i). Here, the derivatives $\widetilde{\lambda}_{hji}(t)$ of the mean functions are known as the (marginal) rate functions. Typically, score equations derived from (3) will still be unbiased estimating equations for these marginal rates, see e.g. Cook and Lawless (2007, Chapter III) and Lin et al. (2000) for the special case of recurrent events. Since these estimating equations are no longer likelihood score equations, robust variance estimation is always needed. Furthermore, the right-censoring times must be assumed marginally independent of the multi-state process.

2.5 Example

We first consider the simple model for survival data depicted in Figure 1 with the transient state 0: "alive" and the absorbing state 1: "dead". Figure 2 shows the Nelson-Aalen estimate for the cumulative $0 \to 1$ transition intensity compared to an estimate of the same quantity based on a model with a piecewise constant hazard using ten intervals with cut-points chosen as death time quantiles. The two estimates are in close agreement and show a high mortality rate for the first approximately two years after transplant after which it seems to level off. Table 2 presents estimates from both a Cox and a Poisson regression model. First, models with the time-fixed covariates disease type, age, and graft type are presented and, next, a time-dependent covariate indicating a history of chronic or acute GvHD is added to the model. It is seen that AML patients have a higher mortality than those with ALL, mortality increases with age, and the mortality rate depends on the graft type. This dependence, however, vanishes once the strongly prognostic time-dependent factor, GvHD is accounted for.

Next, we turn to the three-state illness-death model without recovery, see Figure 3. Here, the transient states are 0: "remission", and 1: "relapse" and

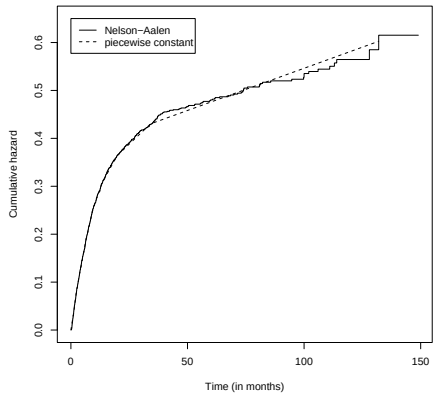


Fig. 2 Nelson-Aalen estimates and estimates based on a piecewise constant hazard for the two-state model for survival data.

Table 2 Cox and Poisson models for the transition intensity in the two-state model.

	Cox			Poisson		
	β	SE	p	β	SE	p
disease (AML vs. ALL)	0.406	0.080	< 0.001	0.409	0.080	< 0.001
age (per 10 years)	0.173	0.026	< 0.001	0.173	0.026	< 0.001
graft type (BM only vs. other)	0.160	0.079	0.043	0.166	0.079	0.035
disease (AML vs. ALL)	0.359	0.080	< 0.001	0.363	0.080	< 0.001
age (per 10 years)	0.156	0.027	< 0.001	0.160	0.027	< 0.001
graft type (BM only vs. other)	0.095	0.079	0.230	0.107	0.079	0.177
GvHD	0.646	0.078	< 0.001	0.529	0.082	< 0.001

the absorbing state is 2: "dead". Figure 4 shows the Nelson-Aalen estimates and corresponding estimates based on models with piecewise constant intensities for transitions out of state 0. As it was the case for the two-state model, non-parametric estimates agree well with those based on piecewise constant intensities. Figure 5a shows the similar estimates for the $1 \rightarrow 2$ transition intensity assuming the process to be Markov, i.e. the transition intensity from state 1 is modelled as depending only on time, t, since transplant. Here, the agreement between the two estimates is less convincing. The intervals in which the intensity is assumed to be constant are the same as previously. However, a closer agreement between the two estimates is obtained by a different choice of intervals, with more cut-points at the beginning of the follow-up time, see Figure 5b.

Table 3 shows estimates in Cox regression models for the three transition intensities still assuming the process to be Markov and only including time-fixed covariates. Age and graft type only seem to affect the direct mortality rate among those in remission and not the relapse rate. In Tables 4 and 5

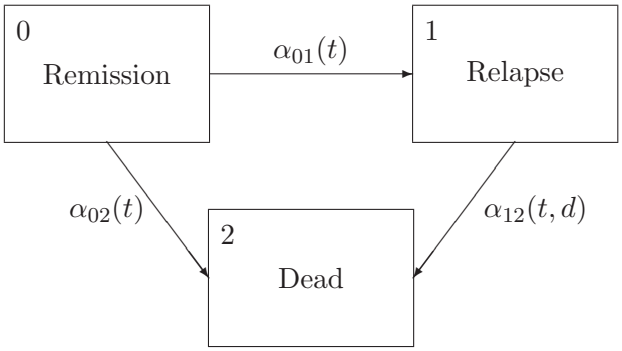


Fig. 3 The illness-death model without recovery.

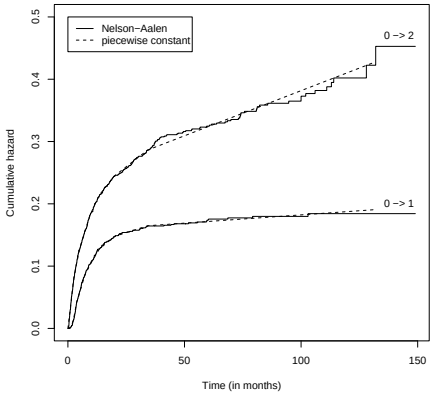
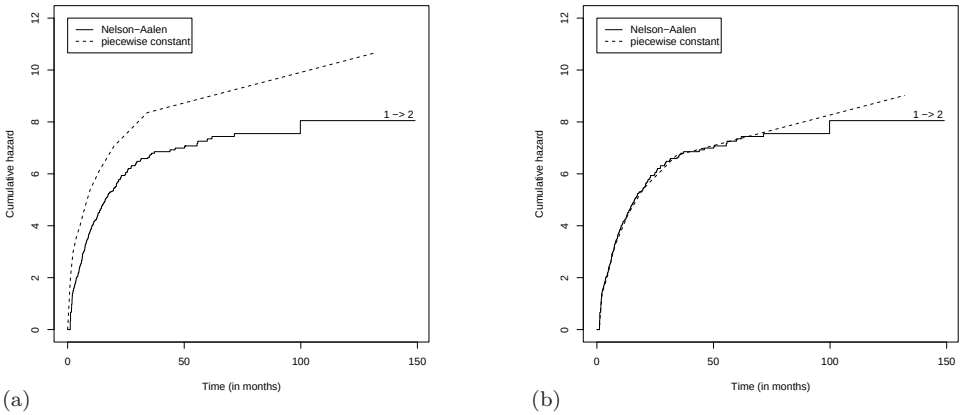


Fig. 4 Nelson-Aalen estimates and estimates based on piecewise constant intensities for transitions out of state 0.

the time-dependent covariate GvHD is also included and this seems to make the effect of graft type disappear both in the Cox model (Table 4) and in the similar Poisson model (Table 5). These two models once again show a very close agreement.

The Markov assumption may be tested by adding time-dependent covariates for the $1 \to 2$ transition intensity depending on the sojourn time spent in state 1 at time t. Including in the Cox model the simple indicator function $I(t-T_{1i} > d_0)$ where d_0 is the median of $t-T_{1i}$ at the $1 \to 2$ transitions gives a coefficient of -1.43, and a p-value less than 0.001. Studying instead a Cox model with time T_{1i} as a covariate gives an estimated coefficient of 0.07, $p < 0.001$.



(a) (b)

Fig. 5 Nelson-Aalen estimates and estimates based on piecewise constant intensities for transitions out of state 1. Both figures use 10 intervals for the piecewise constant model, with case (b) having more cut-points at the very beginning of the follow-up time.

Table 3 Cox models for transition intensities in the illness-death model.

	$0 \to 1$			$0 \to 2$			$1 \to 2$		
	β	SE	p	β	SE	p	β	SE	p
disease (AML vs. ALL)	0.549	0.129	<0.001	0.405	0.098	<0.001	-0.255	0.140	0.068
age (per 10 years)	-0.045	0.044	0.310	0.283	0.032	<0.001	0.002	0.050	0.970
graft type (BM only vs. other)	0.108	0.134	0.420	0.186	0.095	0.051	-0.048	0.173	0.780

Table 4 Cox models for transition intensities in the illness-death model taking into account graft versus host disease as a time-dependent variable (combined acute and chronic - whichever happens first).

	$0 \to 1$			$0 \to 2$			$1 \to 2$		
	β	SE	p	β	SE	p	β	SE	p
disease (AML vs. ALL)	0.563	0.130	<0.001	0.334	0.098	<0.001	-0.271	0.142	0.055
age (per 10 years)	-0.040	0.045	0.370	0.263	0.033	<0.001	-0.002	0.050	0.970
graft type (BM only vs. other)	0.126	0.135	0.350	0.085	0.096	0.370	-0.046	0.172	0.790
GvHD	-0.184	0.134	0.170	1.040	0.098	<0.001	0.103	0.141	0.470

Table 5 Poisson models for transition intensities in the illness-death model taking into account graft versus host disease (combined). The follow-up time is split into 10 intervals.

	$0 \to 1$			$0 \to 2$			$1 \to 2$		
	β	SE	p	β	SE	p	β	SE	p
disease (AML vs. ALL)	0.565	0.130	< 0.001	0.338	0.098	0.001	-0.292	0.141	0.038
age (per 10 years)	-0.039	0.044	0.375	0.263	0.033	< 0.001	-0.007	0.049	0.888
graft type (BM only vs. other)	0.128	0.135	0.342	0.090	0.095	0.346	-0.031	0.171	0.856
GvHD	-0.199	0.134	0.138	1.047	0.098	< 0.001	0.123	0.140	0.380

Finally, fitting a Poisson model with follow-up split according to both time t since BMT and duration, $t-T_1$ in state 1, gives a coefficient of -0.03, $p < 0.001$ for the former and -0.09, $p < 0.001$ for the latter. So, in all cases, the Markov assumption is clearly rejected and it is of interest to study effects of covariates when duration in state 1 is properly accounted for, see Table 6. Here, two Cox models are fitted, one with time, t, since BMT as the baseline time variable and adjusting for duration, $t - T$, using a piecewise constant function (10 intervals), one where the roles of t and $t - T$ are interchanged, and finally a Poisson model where both t and $t - T$ are adjusted for in 10 intervals. Note, as mentioned earlier, that for the Poisson model no baseline time variable is appointed. The change of the estimates is not dramatic though in some of the models disease is, formally, significantly associated with mortality with AML patients now having a slightly smaller intensity. The effect of duration is very marked. This is illustrated in Figure 6 showing the Nelson-Aalen estimate for A_{12} with duration as the time variable.

Table 6 Cox and Poisson models for the $1 \to 2$ transition intensity taking into account both t and $t - T$.

	Cox (time=t)			Cox (time=$t - T$)			Poisson		
	β	SE	p	β	SE	p	β	SE	p
disease (AML vs. ALL)	-0.262	0.141	0.064	-0.260	0.141	0.065	-0.289	0.140	0.039
age (per 10 years)	0.013	0.050	0.790	0.021	0.050	0.670	0.010	0.049	0.846
graft type (BM only vs. other)	-0.034	0.168	0.840	-0.055	0.172	0.750	-0.022	0.167	0.896
GvHD	0.085	0.141	0.550	0.055	0.143	0.700	0.110	0.140	0.430

Note that ignoring what happens after relapse simplifies the model considerably. Doing that we are left with the competing risks model depicted

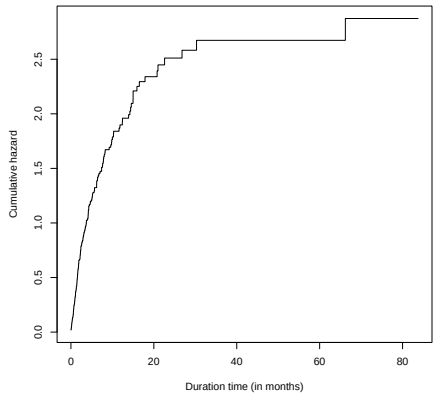


Fig. 6 Nelson-Aalen estimate for cumulative $1 \rightarrow 2$ intensity as a function of duration in state 1.

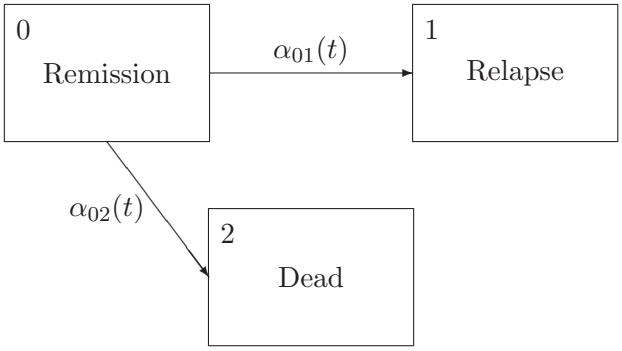


Fig. 7 The competing risks model.

in Figure 7 which now has a single transient state 0: "remission" and two absorbing states, 1: "relapse" and 2: "dead" (in remission). In this model, the transition intensity estimates for $\alpha_{01}(\cdot)$ and $\alpha_{02}(\cdot)$ are the same as those shown in Figure 4 and Tables 3-5. Note that the competing risks model is always Markov.

3 Models for transition and state occupation probabilities

As illustrated in Section 2, intensity-based MSMs are rich and standard software offers the necessary means to perform the analysis. However, since the

interpretation of probabilities is more simple than that of intensities, it is of considerable interest to extend the methods from the previous section with techniques for inference for MSM transition probabilities and state occupation probabilities. For certain MSMs, including Markov processes, explicit formulas relate such probabilities to transition intensities, thereby allowing for simple plug-in probability estimation once intensity models are established. We shall first (Section 3.1) review techniques of that kind.

For regression situations, however, plug-in methods do not provide us with simple parameters describing the association between covariates and outcome probabilities. This is because of the non-linearity of the relation between intensities and probabilities and, hence, even intensity models with a simple link function (such as the Cox model or the additive hazard model), lead to complicated relations between covariates and outcome probabilities. For these reasons, direct (marginal) regression models for outcome probabilities are of interest and we shall review a number of such techniques in Section 3.2, including methods based on pseudo-observations and direct binomial regression models.

3.1 Plug-in models based on intensities

Markov processes

Suppose that the multi-state process $X(t), t \in \mathcal{T}$ is Markov and let $\mathbf{P}(s,t)$ be the $(p+1) \times (p+1)$ transition probability matrix, i.e.

$$P_{hj}(s,t) = \text{Prob}(X(t) = j \mid X(s) = h), h, j \in \mathcal{S}, s \leq t.$$

Let $\mathbf{A}(t)$ be the corresponding $(p+1) \times (p+1)$ cumulative transition intensity matrix, i.e.

$$\mathrm{d}A_{hj}(t) = \text{Prob}(X(t + \mathrm{d}t) = j \mid X(t) = h), h \neq j, h, j \in \mathcal{S}$$

and let $A_{hh}(t) = -\sum_{j \neq h} A_{hj}(t)$, that is $\sum_j A_{hj}(t) = 0$.

We assume that $\mathbf{A}$ is absolutely continuous, that is, $\mathrm{d}A_{hj}(t) = \alpha_{hj}(t)\mathrm{d}t$ where

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P_{hj}(t, t + \Delta t)$$

is the $h \to j$ transition intensity. For given $\mathbf{A}$, the transition probability matrix $\mathbf{P}$ is the unique solution to $\mathbf{P}(s,s) = \mathbf{I}$, the $(p+1) \times (p+1)$ identity matrix, and the Kolmogorov forward differential equations

$$\frac{\partial}{\partial t}\mathbf{P}(s,t) = \mathbf{P}(s,t)\boldsymbol{\alpha}(t)$$

or, written coordinate-wise

$$\frac{\partial}{\partial t} P_{hj}(s,t) = \sum_l P_{hl}(s,t)\alpha_{lj}(t).$$

One can show that the solution is the *matrix product-integral*

$$\mathbf{P}(s,t) = \prod_s^t (\mathbf{I} + \boldsymbol{\alpha}(u)\mathrm{d}u)$$

defined by

$$\lim_{\max|s_i - s_{i-1}|\to 0} \prod (\mathbf{I} + \mathbf{A}(s_i) - \mathbf{A}(s_{i-1})),$$

where $s = s_1 < \ldots < s_{i-1} < s_i < \ldots = t$ is a partition of the interval from s to t (e.g. Andersen et al. 1993, Section II.6).

This provides us with the solution to the non-parametric estimation problem for Markov transition probabilities: estimate $A_{hj}(t)$ by the Nelson-Aalen estimator $\widehat{A}_{hj}(t)$, cf. (4), and let $\widehat{A}_{hh} = -\sum_j \widehat{A}_{hj}$. Transition probabilities are then estimated by the Aalen-Johansen estimator, the finite matrix product obtained by plugging the Nelson-Aalen estimator into the product-integral

$$\widehat{\mathbf{P}}(s,t) = \prod_{(s,t]} (\mathbf{I} + \mathrm{d}\widehat{\mathbf{A}}(u)).$$

For the special case of the two-state model for survival data the latter simply gives the Kaplan-Meier estimator for $S(t) = P_{00}(0,t)$.

Large sample properties may be derived, as follows. We have the martingale representation:

$$\widehat{A}_{hj}(t) - \int_0^t J_h(u)\alpha_{hj}(u)\mathrm{d}u = \int_0^t \frac{J_h(u)}{Y_h(u)} \mathrm{d}M_{hj}(u)$$

where M_{hj} is the counting process martingale for N_{hj} and $J_h(t) = I(Y_h(t) > 0)$. By this representation large sample properties for $\widehat{A}_{hj}(t)$ follow from the martingale Central Limit Theorem; this includes consistency, asymptotic normality, standard errors etc. Large sample properties for the Aalen-Johansen estimator $\widehat{\mathbf{P}}(s,t)$ then follow from those of the Nelson-Aalen estimator and (compact) differentiability of the product-integral by the functional delta-method (e.g. Andersen et al. 1993, Section II.8)..

Parametric models with constant transition intensities may also be handled quite easily. This is because, in this case, there exists a simple exponential representation of the product-integral and, thereby, a simple formula for $P_{hj}(s,t)$. For $d = t - s$ this is given by

$$\mathbf{P}(d) = \exp(d\mathbf{A}) = \mathbf{V}\mathrm{diag}(e^{\boldsymbol{\rho} d})\mathbf{V}^{-1} \tag{6}$$

where $\boldsymbol{\rho}$ are the eigenvalues for the intensity matrix, $\mathbf{A}$, and $\mathbf{V}$ the matrix of eigenvectors, that is, $\mathbf{A}$ is estimated by maximum likelihood and plugged into (6).

The model with piecewise constant intensities can also be handled though the expressions for transition probabilities become slightly more involved. Suppose we wish to estimate $P_{hj}(s,t)$ where s and t belong to adjacent intervals, $[\tau_{\ell-1}, \tau_\ell)$ and $[\tau_\ell, \tau_{\ell+1})$ in which the intensities are constant. Then we may use

the exponential formulas from (6) for $P_{hm}(s,\tau_\ell)$ and $P_{mj}(\tau_\ell,t)$, $m \in \mathcal{S}$ and the *Chapman-Kolmogorov equations*:

$$P_{hj}(s,t) = \sum_{m\in\mathcal{S}} P_{hm}(s,\tau_\ell)P_{mj}(\tau_\ell,t)$$

to estimate $P_{hj}(s,t)$.

For the models with constant or piecewise constant intensities, standard errors for the transition probability estimates are derived from the likelihood-based estimated covariance matrix for $\widehat{\alpha}_{hj\ell}$ and the delta-method.

For Markov regression models with time-fixed covariates (both for the semi-parametric Cox model (1) and for the similar Poisson model) transition probabilities for given covariates, Z_0,

$$P_{hj}(s,t \mid Z_0)$$

may be estimated completely analogously by plugging the estimated regression intensities into the product-integral. Thereby, such probabilities may be predicted for given covariates and standard errors for the predictions may be obtained via the delta-method, Andersen et al. (1991) and Shu and Klein (2005) for the Cox model and the additive model, respectively, in general Markov processes, and Cheng et al. (1998), Shen and Cheng (1999) and Scheike and Zhang (2003) for the competing risks model with Cox hazards, additive hazards and more flexible hazards, respectively. However, as mentioned above, this does not lead to simple relations between covariates and transition probabilities. As an example we study the Markov illness-death model (Figure 3) with Cox type $0 \to 1, 0 \to 2$ and $1 \to 2$ transition intensities. In this model, the transition probability $P_{01}(0,t)$ is given by

$$P_{01}(0,t \mid Z) = \int_0^t P_{00}(0,u- \mid Z)\alpha_{01}(u \mid Z)P_{11}(u+,t \mid Z)\mathrm{d}u \qquad (7)$$

with

$$P_{00}(0,u \mid Z) = \exp\left(-A_{01}(u \mid Z) - A_{02}(u \mid Z)\right)$$

and

$$P_{11}(u,t \mid Z) = \exp\left(-\int_u^t \alpha_{12}(x \mid Z)\mathrm{d}x\right).$$

Thus, for $\alpha_{hj}(u \mid Z_i) = \alpha_{hj0}(u)\exp(\mathrm{LP}_{hji})$ the way in which $P_{01}(0,t \mid Z)$ depends on Z is not described by simple parameters.

A note on interval-censoring

In the derivation of the likelihood, (3), continuous observation of $X(\cdot)$ was assumed, i.e. times of transitions were observed exactly, except possibly for right-censoring. For a Markov model with piecewise constant intensities the

likelihood for interval-censored data may be written down in terms of intensities based on the explicit relation between transition intensities and probabilities. For a single individual, observed at times $s_0, s_1, \ldots, s_r$ to be in states $x_0 = X(s_0), x_1 = X(s_1), \ldots, x_r = X(s_r)$ the likelihood contribution is

$$\prod_{j=1}^{r} P_{x_{j-1}x_j}(s_{j-1}, s_j).$$

This means that Markov models with piecewise constant transition intensities (including regression models) may be handled quite easily, e.g. Kay (1986). Non-parametric inference based on interval-censored data have been studied in some special cases, including the 3-state Markov illness-death model without recovery by Frydman (1992, 1995), see also Commenges (2002).

Non-Markov processes

For semi-Markov processes without loops, that is, when only a finite number of paths from any state $h \in \mathcal{S}$ to another state $j \in \mathcal{S}$ is possible, explicit expressions for transition probabilities like (7) are available. As a first example we consider the semi-Markov illness-death model without recovery and where the transition intensity from 1 to 2, $\alpha_{12}(t, t - T_1 \mid Z)$ depends on both "calendar" time, t, and duration $t - T_1$ in state 1, where T_1 is the time of transition from 0 to 1. Here, $P_{01}(0, t)$ is still given by (7) with P_{11} now specified as

$$P_{11}(u, t \mid Z, T_1) = P_{11}(u, t \mid Z, u) = \exp\left(-\int_u^t \alpha_{12}(x, x - u \mid Z)\mathrm{d}x\right).$$

Asymptotics for this model was studied by Shu et al. (2007) for the special case of $\alpha_{12}(x, x - u \mid Z) = \alpha_{12}(x - u \mid Z)$ (i.e., a homogeneous semi-Markov as described in Section 2.1).

Also in a more complicated semi-Markov model with an extra transient state representing, e.g. GvHD, explicit expressions for the transition probabilities are available, see Figure 8. Thus, we can write the probability $P_{02}(0, t)$ as

$$P_{02}(0, t) = Q_{02}(0, t) + Q_{012}(0, t),$$

the sum of the probabilities of the paths $0 \to 2$ and $0 \to 1 \to 2$, respectively. Here,

$$Q_{02}(0, t) = \int_0^t P_{00}(0, u-)\alpha_{02}(u)P_{22}(u+, t \mid u)\mathrm{d}u$$

and

$$Q_{012}(0, t) = \int_0^t P_{00}(0, u-)\alpha_{01}(u) \int_u^t P_{11}(u+, x- \mid u)\alpha_{12}(x, x-u)P_{22}(x+, t \mid x)\mathrm{d}x\mathrm{d}u$$

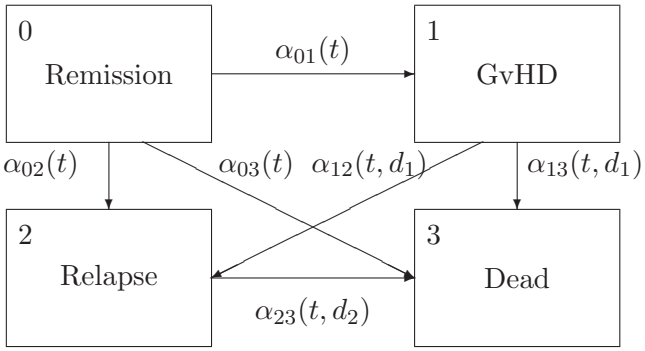


Fig. 8 An extended model for BMT.

with

$$P_{11}(s,t \mid T_1) = \exp\left(-\int_s^t (\alpha_{12}(u, u-T_1) + \alpha_{13}(u, u-T_1))\mathrm{d}u\right)$$

and

$$P_{22}(s,t \mid T_2) = \exp\left(-\int_s^t \alpha_{23}(u, u-T_2)\mathrm{d}u\right).$$

State occupation probabilities

For general multi-state models, Datta and Satten (2001) studied estimation of the state occupation probabilities $\pi_h(t), h \in \mathcal{S}$. They showed that the product-integral estimator is consistent without the Markov assumption and related this fact to the estimation of marginal rates as discussed in Section 2.

3.2 Direct models for probabilities

In some models, transition probabilities may be estimated directly. Thus, for a general non-Markov illness-death process without recovery, Meira-Machado et al. (2006) derived estimators for the transition probabilities $P_{00}(s,t)$, $P_{01}(s,t)$, $P_{11}(s,t)$, as follows. If T_0 and T_2 represent the sojourn time spent in state 0 and the time to absorption in state 2, respectively, and if H is the survival function for T_0 then

$$P_{00}(s,t) = \frac{H(t)}{H(s)}, P_{01}(s,t) = \frac{\mathrm{E}(\phi_{st}(T_0,T_2))}{H(s)}, P_{11}(s,t) = \frac{\mathrm{E}(\widetilde{\phi}_{st}(T_0,T_2))}{\mathrm{E}(\widetilde{\phi}_{ss}(T_0,T_2))},$$

where $\phi_{st}(u,v) = I(s < u \leq t, v > t)$ and $\widetilde{\phi}_{st}(u,v) = I(u \leq s, v > t)$. Here, H can be estimated by the Kaplan-Meier estimator, $\widehat{H}$, while the expectations $\mathrm{E}(\phi_{st}(T_0, T_2))$ and $\mathrm{E}(\widetilde{\phi}_{st}(T_0, T_2))$ can be estimated by Kaplan-Meier integrals of the form

$$\sum_{i=1}^{n} w_i \phi_{st}(\widetilde{T}_{0i}, \widetilde{T}_{2i})$$

where $\widetilde{T}_{0i}, \widetilde{T}_{2i}$ are the, possibly right-censored, observations of T_0, T_2 for individual i and w_i the "Kaplan-Meier weight" associated with $\widetilde{T}_{2i}$, i.e.

$$w_i = \frac{\mathrm{d}N_{02}(\widetilde{T}_{2i}) + \mathrm{d}N_{12}(\widetilde{T}_{2i})}{Y_0(\widetilde{T}_{2i}) + Y_1(\widetilde{T}_{2i})} \widehat{S}(\widetilde{T}_{2i}-).$$

Here, S is the survival function for T_2 and $\widehat{S}$ its Kaplan-Meier estimator. Note that, without right-censoring, the estimator of $P_{hj}(s,t)$ reduces to the relative frequency of processes in state j at time t among those in state h at time $s < t$. Meira-Machado et al. (2006) derived large sample properties of these estimators which may be generalized to more complicated non-Markov processes.

Furthermore, for a transient state the state occupation probability may be estimated by "Kaplan-Meier differences" (Pepe 1991). As a simple example, let us once more study the illness-death model without recovery where T_0 is the time spent in state 0 and T_2 the time to death with survival functions H and S, respectively. Then $H(t)$ is the probability that the process is in state 0 at time t and $S(t)$ is the probability that it is in either state 0 or state 1 at time t. With Kaplan-Meier estimators $\widehat{H}$ and $\widehat{S}$ for T_0 and T_2, respectively, the state occupation probability $\pi_1(t)$ for the transient state 1 can, therefore, be estimated without a Markov assumption by

$$\widehat{\pi}_1(t) = \widehat{S}(t) - \widehat{H}(t).$$

This approach may be used for transient states in more general MSMs.

Regression models

Without censoring, state occupation indicators $I(X_i(t) = h)$ would always be observed and could thereby be used as outcome variables in generalized linear models for

$$\pi_{hi}(t \mid Z_i) = \mathrm{E}(I(X_i(t) = h \mid Z_i)),$$

that is model of the form

$$g(\pi_{hi}(t)) = \mathrm{LP}_{hi}(t)$$

with link function g. Here, the estimating equations would be $\sum_i U_i(\beta, t) = 0$, for (all or) selected $t-$values, with

$$U_i(\beta, t) = \left(\frac{\partial}{\partial \beta} g^{-1}(\mathrm{LP}_{hi}(t))\right)^{\mathsf{T}} \mathbf{V}_i^{-1}(I(X_i(t) = h) - g^{-1}(\mathrm{LP}_{hi}(t))). \qquad (8)$$

In (8), $\mathbf{V}_i$ is a working covariance matrix, frequently chosen simply to be the identity.

With censoring, $I(X_i(t) = h)$ is not always observed and modifications to (8) are needed. One possibility is to replace $I(X_i(t) = h)$ by its pseudo-observation (e.g. Andersen et al. 1993; Andersen and Klein 2007) given by

$$\widehat{\pi}_{hi}(t) = n\widehat{\pi}_h(t) - (n-1)\widehat{\pi}_h^{-i}(t)$$

where $\widehat{\pi}_h(t)$ is a well-behaved estimator for $\pi_h(t)$ based on the entire sample of size n while $\widehat{\pi}_h^{-i}(t)$ is the same estimator applied to the sample of size $n-1$ obtained by eliminating subject i. For this approach to work, right-censoring should be independent of both the multi-state process $X_i(\cdot)$ and of the covariates, Z_i.

An alternative method, based on inverse probability of censoring weighting was studied by Scheike and Zhang (2007). Here, the starting point is once more (8) where $I(X_i(t) = h)$ is now replaced by

$$\frac{I(X_i(t) = h)I(C_i > t)}{G_C(t)}$$

with G_C denoting (an estimate of) the censoring distribution, $G_C(t) = \text{Prob}(C_i > t)$. For the special case of the competing risks model this approach was shown by Graw et al. (2008) to be equivalent to that based on pseudo-observations. For this model, both methods are also closely related to the regression techniques suggested by Fine and Gray (1999)(see also Fine 2001) for the cumulative incidence function, $P_{0h}(0,t)$ that was studied via the sub-distribution hazard

$$\widetilde{\alpha}_h(t) = \frac{\partial}{\partial t}\log(-\log(1 - P_{0h}(0,t))).$$

The Cox regression score equations were modified by inverse probability of censoring weights. We prefer to formulate the models directly in terms of $P_{0h}(0,t)$, mainly because of the awkward interpretation of the sub-distribution hazard, which is

$$\widetilde{\alpha}_h(t)\mathrm{d}t = \text{Prob}(\text{failure from cause } h \text{ in } (t, t+\mathrm{d}t) \mid$$

either alive at time $t-$ *or* failure from a competing cause in $[0,t))$.

3.3 Example

We first study the simple two-state model for mortality, Figure 1. Figure 9 shows the Kaplan-Meier estimator for the survival probability, $S(t) = P_{00}(0,t)$, compared to an estimate based on a model with a hazard assumed to be constant in each of ten intervals. The agreement is very close. Estimates for given covariates can be obtained by plugging-in estimated baseline hazards

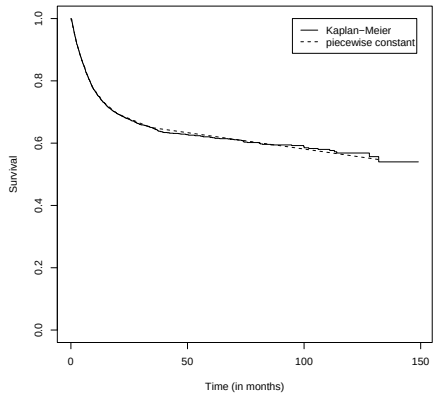


Fig. 9 Kaplan-Meier estimate and estimate based on a piecewise constant hazard in the two-state model.

and coefficients from Table 2. Because of the one-to-one relationship between the transition intensity and the survival probability, differences between such predicted curves for different covariate patterns are directly reflected by the regression coefficients.

For the competing risks model, Figure 7, the estimated cumulative incidences, $\widehat{P}_{01}(0,t)$ and $\widehat{P}_{02}(0,t)$ for relapse and death in remission, respectively, are shown in Figure 10. Since $\pi_0(0) = 1$ the cumulative incidences equal the state occupation probabilities. Both the non-parametric Aalen-Johansen estimates and those based on piecewise constant cause-specific hazards are shown and are seen to be in close agreement. Figure 11 shows the Aalen-Johansen (in fact, Kaplan-Meier) estimate for the leukemia-free survival function, $P_{00}(0,t)$ in the competing risks model. Note that $P_{00} + P_{01} + P_{02} = 1$.

Again, plug-in estimation of the cumulative incidences $P_{01}(0,t)$ and $P_{02}(0,t)$ is possible, see Figure 12 where this is done for an AML patient with average age (32 years) and graft type BM. However, differences between such curves for different covariate patterns are not simple functions of estimates like those presented in Table 3 and direct regression modelling of the cumulative incidences is of interest. Table 7 shows estimates in such models based on either the estimation procedure suggested by Fine and Gray (1999) or based on pseudo-observations. The latter is done for both the cloglog link function (like in the Fine-Gray procedure) and for the logit link in both cases using 3 time points in the estimating equations. These were chosen of quartiles for either the death or the relapse times. The parameter estimates here reflect the direct association between covariates and the (transformed) cumulative incidence functions. As also noted by Klein and Andersen (2005), estimates based on pseudo-observations with a cloglog link are close to those based on the Fine and Gray estimation technique while the latter have smaller standard errors.

Table 7 Estimates in direct regression models for cumulative incidences.

	Fine & Gray			pseudo - cloglog			pseudo - logit		
	β	SE	p	β	SE	p	β	SE	p
relapse									
disease	0.472	0.131	<0.001	0.394	0.147	0.007	0.415	0.155	0.008
age	-0.093	0.045	0.040	-0.089	0.050	0.077	-0.095	0.053	0.074
graft type	0.057	0.125	0.650	0.093	0.140	0.509	0.098	0.148	0.509
death									
disease	0.337	0.097	0.001	0.363	0.121	0.003	0.409	0.134	0.002
age	0.282	0.032	<0.001	0.268	0.040	<0.001	0.297	0.045	<0.001
graft type	-0.085	0.090	0.343	-0.100	0.121	0.408	-0.111	0.134	0.409

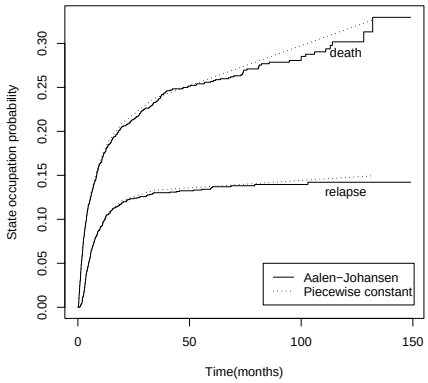


Fig. 10 Cumulative incidence estimates (state occupation probabilities): Aalen-Johansen and piecewise constant hazards.

Note that age, probably because of its strong positive association with the mortality rate $\alpha_{02}(\cdot)$, Tables 3-5, is now negatively associated with $P_{01}(0,t)$. Graft type has no strong association with any of the cumulative incidences. Results for logit and cloglog links lead to the same qualitative conclusions. Note that, since these are models for the *cumulative probabilities* $P_{01}(0,t)$ and $P_{02}(0,t)$, the time-dependent covariate GvHD should not be included in the model. This is analogous to ordinary survival analysis where the cumulative survival probability cannot be calculated from a hazard model with time-dependent covariates

For the three-state illness-death model without recovery, Figure 13a shows the Aalen-Johansen estimates for the probabilities $P_{01}(0,t)$, $P_{02}(0,t)$, and $P_{12}(m,t)$ where $m = 6.22$ months is the median relapse time. Figure 13b shows the predicted values of $P_{01}(0,t)$ and $P_{02}(0,t)$ using the plug-in estimator for two individuals that differ in age.

Since the analyses in Section 2 showed that the Markov assumption underlying these estimates was highly questionable, both time t and duration $t-T$

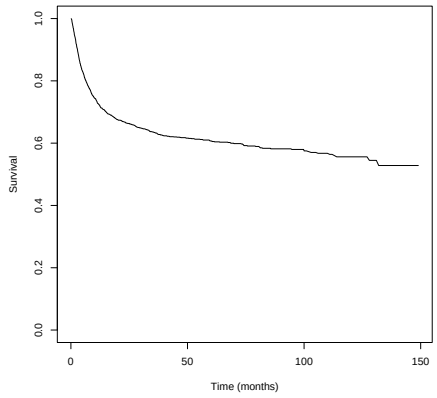


Fig. 11 Aalen-Johansen (Kaplan-Meier) estimate for leukemia-free survival function.

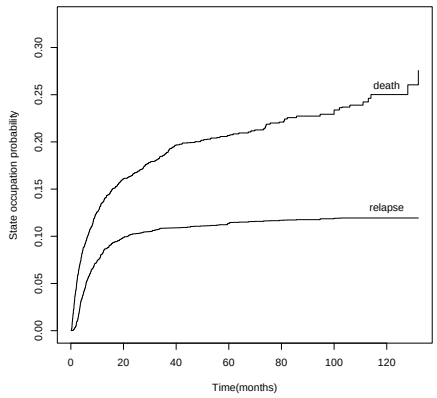
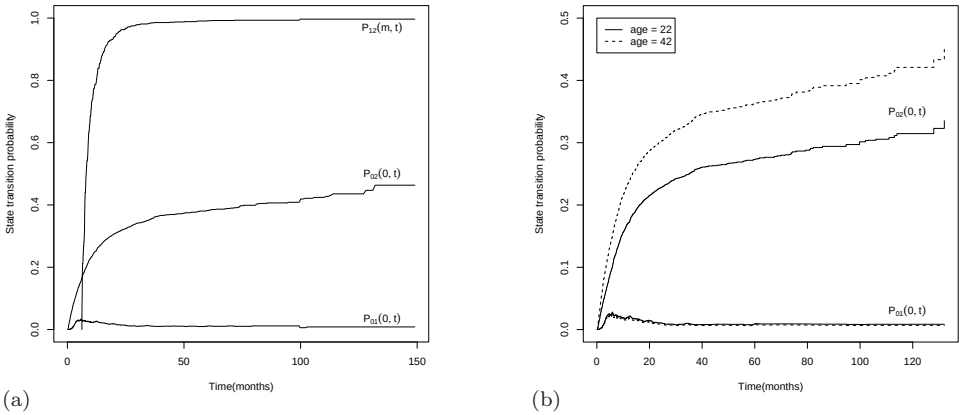


Fig. 12 Cumulative incidence (state occupation probability) estimates for AML patient, 32 years, BM only.

should be used when estimating the transition probability $P_{12}(s,t)$ using the plug-in model. In Figure 14 this is done using the simple semi-Markov model with a (log-)linear effect of duration in state 1. The curves have the same general shape with an early steep increase.

Another way of relaxing the Markov assumption is to restrict attention to state occupation probabilities, $\pi_h(t)$. Figure 15 compares the Aalen-Johansen estimator to the Pepe "Kaplan-Meier difference estimator" for $\pi_1(t)$. As expected, the curves are close.



(a) (b)

Fig. 13 Estimated transition probabilities in a three-state Markov illness-death model: (a) Aalen-Johansen estimator of $P_{01}(0,t)$, $P_{02}(0,t)$, and $P_{12}(6.22,t)$; (b) The predicted probabilities $P_{01}(0,t)$ and $P_{02}(0,t)$ under the Cox model comparing two patients with a 20-year age difference (both AML and BM only).

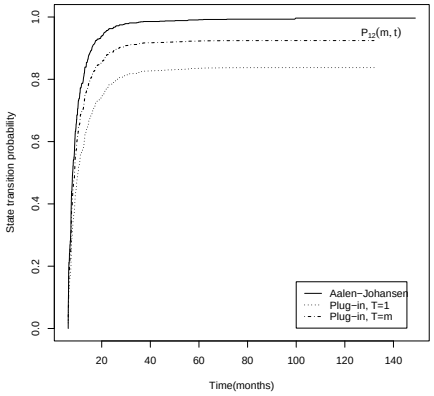


Fig. 14 The plug-in model for the transition probability $P_{12}(6.22, t \mid T)$ for $T = 6.22$ and $T = 1$ compared to the Aalen-Johansen estimator (assuming a Markov illness-death model).

Finally, Figure 16 compares the Aalen-Johansen estimator to the Meira-Machado estimators and, again, the curves are rather close. This was to be expected for the state occupation probabilities, $P_{01}(0,t) = \pi_1(t)$ and $P_{02}(0,t) = \pi_2(t)$, but it seems to be the case also for the "genuine" transition probability $P_{12}(m,t)$ in spite of the lack of fit of the Markov model.

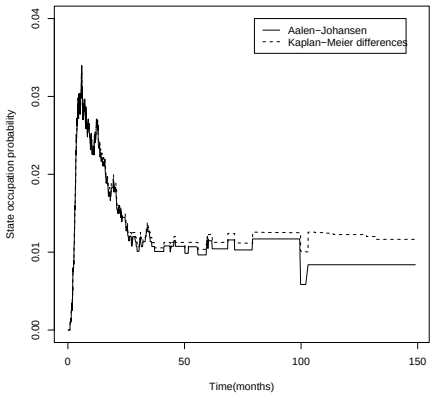
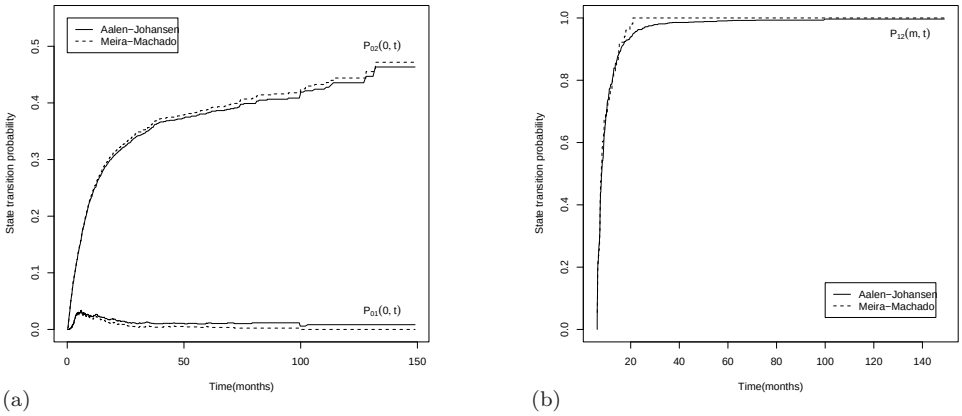


Fig. 15 The Aalen-Johansen estimator and the Pepe Kaplan-Meier difference estimator for the state occupation probability $\pi_1(t)$ in a three-state illness-death model.



(a) (b)

Fig. 16 Estimated transition probabilities in a three-state illness-death model using the Meira-Machado estimators and the Aalen-Johansen estimators: (a) $P_{01}(0,t)$, $P_{02}(0,t)$; (b) $P_{12}(6.22,t)$.

4 Comments

In this review we have presented a series of statistical methods for MSMs which may be useful in the analysis of follow-up data for patients with BMT. One class of models was based on transition intensities. These are the most fundamental parameters in MSMs and they are the parameters which enter directly in the likelihood for continuously observed follow-data, cf. (3). We focussed on two broad classes of intensity models: non- or semi-parametric models, and parametric models with piecewise constant transition intensities. One purpose

of our illustrative example was to emphasize that these models, in fact, tend to provide very similar results. However, for the model with piecewise constant intensities the choice of intervals could affect the results, see, e.g. Figure 5, and this choice may not always be obvious.

Another purpose of our illustrative example was to demonstrate that the MSMs may provide insight into the data which may be overlooked if one instead analyzes the time to death using more simple survival analysis techniques. Thus, prognostic factors for survival may influence different transition intensities quite differently. As an example, the time-dependent covariate GvHD which was strongly associated with the overall mortality, Table 2, seemed only to increase the intensity of death in remission but not the relapse intensity or the death intensity after relapse, Tables 4-6.

Another class of models focussed on outcome probabilities, that is, state occupation probabilies and transition probabilities. One advantage of such models is the more direct interpretation of probabilities than intensities. While transition intensities provide a *local* (in time) description of the dynamics of the model, the probabilities give a global description which has been accumulated over time. For some intensity-based models (Markovian models and semi-Markov models without loops) transition probabilities were easily estimable by plug-in methods while the (marginal) state occupation probabilities could be estimated in more general classes of multi-state models (Figures 9-14). When probabilities were estimated by plugging-in regression models for transition intensities no simple relationship between covariates and probabilities was obtained though predictions for given covariates were quite simple. In such situations, direct regression models for the outcome probabilities provided an alternative option which we exemplified using pseudo-observations, Table 7.

Some limitations of the methods should be mentioned. First of all, we have only exemplified analysis of continuously observed data where all transition times were observed exactly. For interval-censored data, the only general approach (for Markovian models) seems to be models with piecewise constant intensities. In this connection it is reassuring to note the similarity between results from models with piecewise constant intensities and those from non- or semi-parametric inference. One should notice that the complexity of the inference increases with the number of possible transitions in the model, and we have only exemplified analyses with few states. An obvious extension of the models that we have discussed for our example data is the extended model with an additional state for GvHD, Figure 8, either combining acute and chronic GvHD, as we did when treating it as a time-dependent covariate, or using separate states for the two. However, due to lack of space we did not include analyses of such an extended model in our examples.

One final remark is that our example should be regarded as purely illustrative with the purpose of showing how the different models we have discussed

may be handled in practice. Thus, our example was not intended to provide definitive analyses for the BMT data. For this to be the case, much more attention must be paid to the goodness of fit of the models. For models based on intensities, techniques known from survival analysis may be applied while goodness of fit of models for pseudo-observations was discussed by Klein and Andersen (2005).

Acknowledgements The research was supported by grant R01-54706-12 from the National Cancer Institute and Danish Natural Science Research Council, grant 272-06-0442 "Point process modelling and statistical inference". We are grateful to CIBMTR for providing us with the example data. An early version of this paper was presented at a workshop held in Lillehammer, Norway, sponsored by the Centre for Advanced Study at the Norwegian Academy of Sciences and Letters.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P. K., Hansen, L. S., and Keiding, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process. *Scandinavian Journal of Statistics*, 18:153–167.

Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11:91–115.

Andersen, P. K. and Klein, J. P. (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics*, 34:3–16.

Cheng, S. C., Fine, J. P., and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, 54:219–228.

Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*, 5:315–327.

Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical methods in Medical Research*, 11:167–182.

Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. Springer-Verlag, New York.

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

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, 2:85–97.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94:496–509.

Frydman, H. (1992). 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*, 54:853–866.

Frydman, H. (1995). Nonparametric estimation of a Markov 'illness-death' process from interval- censored observations, with application to diabetes survival data. *Biometrika*, 82:773–789.

Graw, F., Gerds, T. A., and Schumacher, M. (2008). On pseudo-values for regression analysis in multi-state models. *manuscript*

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis*, 5:239–264.

Kay, R. A. (1986). Markov model for analysing cancer markers and disease states in survival models. *Biometrics*, 42:855–865.

Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, 61:223–229.

Klein, J. P. and Shu, Y. (2002). Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research*, 11:117–139.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. L. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society — Series B*, 62:711–730.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer-Verlag, New York.

Meira-Machado, L., Ũna-Álvarez, J., and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illnessdeath model. *Lifetime Data Analysis*, 12:325–344.

Meira-Machado, L., Ũna-Álvarez, J., Cadarso-Suárez, C. and Andersen, P. K. (2009). Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research*, 18:to appear.

Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*, 86:770–778.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26:2389–2430.

Scheike, T. H. and Zhang, M. J. (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics*, 59:1036–1045.

Scheike, T. H. and Zhang, M. J. (2007). Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics*, 34:17–32.

Shen, Y. and Cheng, S. C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics*, 55:1093–1100.

Shu, Y. and Klein, J. P. (2005). Additive hazards Markov regression models illustrated with bone marrow transplant data. *Biometrika*, 92:283–301.

Shu, Y., Klein, J. P., and Zhang, M. J. (2007). Asymptotic theory for the Cox semi-Markov illness-death model. *Lifetime Data Analysis*, 13:91–117.