STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:2389–2430
Published online 10 October 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2712

Tutorial in biostatistics: Competing risks and multi-state models

H. Putter[1,*,†,‡], M. Fiocco[1] and R. B. Geskus[1,2,3,‡]

[1]*Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands*
[2]*Amsterdam Health Service, Amsterdam, The Netherlands*
[3]*Academic Medical Center, Amsterdam, The Netherlands*

SUMMARY

Standard survival data measure the time span from some time origin until the occurrence of one type of event. If several types of events occur, a model describing progression to each of these competing risks is needed. Multi-state models generalize competing risks models by also describing transitions to intermediate events. Methods to analyze such models have been developed over the last two decades. Fortunately, most of the analyzes can be performed within the standard statistical packages, but may require some extra effort with respect to data preparation and programming. This tutorial aims to review statistical methods for the analysis of competing risks and multi-state models. Although some conceptual issues are covered, the emphasis is on practical issues like data preparation, estimation of the effect of covariates, and estimation of cumulative incidence functions and state and transition probabilities. Examples of analysis with standard software are shown. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: competing risks; multi-state model; survival analysis; prognostic factors; prediction

1. INTRODUCTION

Standard survival data measure the time span from some time origin until the occurrence of the event of interest. Examples from medical and epidemiological research include the time to leukaemia relapse after bone marrow transplantation and the time from infection by the HIV virus until the development of AIDS. Typically, in medical research survival data are obtained from clinical trials in which the effect of an intervention (treatment) is measured, whereas in epidemiological research data are obtained from observational studies such as cohort studies.

*Correspondence to: H. Putter, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, P.O. Box 9604, 2300 RC, Leiden, The Netherlands.
†E-mail: h.putter@lumc.nl
‡Contributed equally to this tutorial.

Contract/grant sponsor: Zon-MW; contract/grant number: 0032-4633-2324

Received 26 October 2005
Accepted 31 July 2006

In the disease/recovery process, often more than one type of event plays a role. Usually, one type of event can be singled out as the event of interest. The other event types may prevent the event of interest from occurring. Leukaemia relapse or AIDS may be unobservable because the person died before the diagnosis of these events. Caution is needed in estimating the probability of the event of interest occurring in the presence of these so-called competing risks. Treating the events of the competing causes as censored observations will lead to a bias in the Kaplan–Meier estimate if one of the fundamental assumptions underlying the Kaplan–Meier estimator is violated: the assumption of independence of the time to event and the censoring distributions. The Cox proportional hazards model can still be used, but the interpretation of the results is different. This will be outlined in some detail in Section 3.

In other situations, another event may substantially change the risk of the event of interest to occur. If one is only interested in the event of interest *as a first event*, the other event can still be seen as competing. Often, one is also interested in what happens after the first non-fatal event. Then intermediate event types provide more detailed information on the disease/recovery process and allow for more precision in predicting the prognosis of patients. For a leukaemia patient, if the event of interest is death, then relapse becomes an intermediate event worth modelling and not preventing death. Such non-fatal events during the disease course can be seen as *transitions* from one state to another. The time origin is characterized by a transition into an initial, transient, state, such as the start of treatment; the endpoint is an 'absorbing' final transition. Instead of survival data or time-to-event data, data on the *history of events* is available. Multi-state models provide a framework that allow for the analysis of such event history data. They are an extension of competing risk models, since they extend the analysis to what happens after the first event. Multi-state models are the subject of Section 4.

Several of the ideas presented in the sections on competing risks and multi-state models can also be found in Reference [1]. For more information on competing risks and multi-state models we refer to the relevant chapters in the textbooks [2–7]. A recent issue of Statistical Methods in Medical Research, entirely devoted to multi-state models, is also of interest, see e.g. References [1, 8, 9].

This tutorial reviews statistical methods for the analysis of competing risks and multi-state models. Fortunately, the theory that has been developed over the past two decades for the analysis of right censored survival data can be applied to competing risks and multi-state models as well and often most of the analyzes can be performed within the standard statistical packages, but may require some extra effort with respect to data preparation and programming. Section 2 introduces background and notation needed for the sequel of the paper and discusses the implications of the (lack of) independence between the censoring and time-to-event distributions. Sections 3 and 4 discuss competing risks and multi-state models respectively. Each of these sections is concluded with a subsection on available software. We illustrate estimation and modelling aspects of competing risks and multi-state models using the statistical package `R` [10]. The full code for the analyzes performed in this tutorial as well as the data used are available at `http://www.msbi.nl/multistate`.

2. BACKGROUND AND NOTATION

The central role played by time brings about special characteristics for survival data. The observation window during which data are collected causes individuals to have part of their disease history unobserved. If the endpoint of interest has not (yet) occurred at the end of the observation window,

DOI: 10.1002/sim

the event time is right censored. The event may occur between two consecutive observation times within the observation window, leading to interval censored data. In cohort studies, there is less control with respect to occurrence of the event that determines the time origin. For example, HIV infection may have occurred before an individual enters a cohort study on AIDS. If this time origin is unknown, it is left censored. Sometimes extra information on the time origin is available, for instance through stored blood samples in case of HIV infection. Such individuals only provide information from the moment of entry until their endpoint of interest. This is called delayed entry or left truncation.

In the sequel, we restrict to data in which all the event times are observed exactly or right censored. Hence, left censored and interval censored data are not discussed. Left truncated data, however, play a major role in multi-state settings. We assume throughout that all failure time distributions are continuous. In the model for right censored data, each individual i is assumed to have an event time t_i and a censoring time c_i. Observed are $x_i = \min(t_i, c_i)$ and $\delta_i = I(t_i \leqslant c_i)$, indicating whether t_i was observed ($\delta_i = 1$) or not ($\delta_i = 0$). The event times and censoring times of the individuals in the data set are seen as a random sample $(X_1, C_1), \ldots, (X_n, C_n)$ from a survival distribution $X_i \sim S$, with $S(t) = \text{Prob}(T > t)$, and a censoring distribution $C_i \sim G$. The basic assumption of the standard models for right censored data is that the censoring distribution and the event time distribution are independent (possibly conditionally on the covariates included in the model). Then, at each point in time, the individuals who are censored can be represented by those who remain under observation. Therefore, the hazard, defined for continuous distributions as

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{\text{Prob}(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t} \tag{1}$$

plays a fundamental role in the analysis of right censored survival data. By the independence assumption, the hazard of the individuals that are censored is equal to the hazard of the individuals that remain in follow-up.

The hazard completely describes the survival distribution. It can be derived from the survival function $S(t)$ through

$$\lambda(t) = \frac{1}{S(t)} \lim_{\Delta t \downarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t} = -\frac{\mathrm{d} \log S(t)}{\mathrm{d}t} \tag{2}$$

The cumulative hazard is defined by

$$\Lambda(t) = \int_0^t \lambda(s)\, \mathrm{d}s \tag{3}$$

The survival function can be found from the cumulative hazard through the relation

$$S(t) = \exp(-\Lambda(t)) \tag{4}$$

It is instructive, particularly in view of the extension in Section 3, to give a heuristic derivation of the Kaplan–Meier estimator of the survival function. Let $0 < t_1 < t_2 < \cdots < t_N$ be the ordered distinct time points at which events occur. Let $R(t)$ be the *risk set* (those subjects that are in follow-up and have not reached their event) at time t. For each t_j, define $R_j = R(t_j)$ to be the risk set at t_j, and n_j the size of this risk set, the *number at risk*. For each t_j, define d_j to be the number of observed events at t_j. Even though our restriction to continuous distributions prevents the occurrence of tied events (i.e. $d_j > 1$), they may occur because of rounding errors.

DOI: 10.1002/sim

The Kaplan–Meier estimator treats the data 'as is', so it assumes that the distribution is discrete instead of continuous, with the events only occurring at these observed time points. Consider the conditional probability of failing at t_j, given still alive just before time t_j. Since events are assumed only to occur at the observed event times, 'alive just before time t_j' is equivalent to 'alive beyond the previous time point t_{j-1}'. In general, 'alive just before time t' is often denoted as 'alive at $t-$'; the distinction between t and $t-$ is only needed if the distribution is discrete. Hence we can write the conditional probability of failing at t_j, given still alive just before time t_j as $\lambda(t_j) = \text{Prob}(T = t_j | T > t_{j-1})$, a discretized form of the hazard function of equation (1). Under the assumption of independent censoring, subjects in the risk set are representative for all subjects alive at t_j-, so $\lambda(t_j)$ can be estimated simply by the at risk sample proportion that fail at t_j, i.e. by

$$\widehat{\lambda}(t_j) = \frac{d_j}{n_j} \tag{5}$$

The probability of surviving up to t_j is the product of the probability of surviving up to t_{j-1} and the conditional probability of surviving up to t_j given still alive beyond t_{j-1}; in formula form

$$\widehat{S}(t_j) = \widehat{S}(t_{j-1})(1 - \widehat{\lambda}(t_j)) = \widehat{S}(t_{j-1})\left(1 - \frac{d_j}{n_j}\right) \tag{6}$$

By repeatedly applying (6) one then finds the Kaplan–Meier estimator

$$\widehat{S}(t) = \prod_{j:t_j \leqslant t} \left(1 - \frac{d_j}{n_j}\right) \tag{7}$$

If the sample size increases, the number of event times increases as well, and the Kaplan–Meier estimate approaches a continuous distribution. Also, the Kaplan–Meier survival estimate and the estimate based on the exponential form of the survival function (using (4) and the estimate of the hazard in the exponential) become similar.

The effect of covariates on disease progression is most often modelled using the Cox proportional hazards model. In its simplest form, the hazard for a subject with covariate values $\mathbf{Z} = (Z_1, \ldots, Z_p)$ is assumed to be

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{Z})$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients and $\lambda_0(t)$ is the baseline hazard. Here and in the sequel, we will use $\boldsymbol{\beta}^\top \mathbf{Z}$ as a short-hand notation for $\sum_{k=1}^{p} \beta_k \times Z_k$. Assuming all event times are distinct, the parameter vector $\boldsymbol{\beta}$ is found by maximising the partial likelihood. This is a product, over the event times, of a quotient that compares the hazard of the individual with the event at t_j to the hazard of all the individuals at risk at t_j:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{N} \frac{\exp(\boldsymbol{\beta}^\top \mathbf{Z}_j)}{\sum_{l \in R_j} \exp(\boldsymbol{\beta}^\top \mathbf{Z}_l)}$$

Note that the baseline hazard cancels out. The estimate $\widehat{\boldsymbol{\beta}}$ is used in Breslow's estimate of the baseline cumulative hazard

$$\widehat{\Lambda}_0(t) = \sum_{j:t_j \leqslant t} \frac{1}{\sum_{l \in R_j} \exp(\widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_l)}$$

DOI: 10.1002/sim

A number of methods exist to deal with tied event times which fall outside the scope of this tutorial.

Sometimes, one may want to allow the baseline hazard to be different across subgroups $h = 1, \ldots, m$, called strata:

$$\lambda_h(t|\mathbf{Z}) = \lambda_{h,0}(t)\exp(\boldsymbol{\beta}^\top \mathbf{Z})$$

Parameter estimation in this stratified Cox model is performed by maximization of the partial likelihood per stratum

$$L(\boldsymbol{\beta}) = \prod_{h=1}^{m} L_h(\boldsymbol{\beta}) \tag{8}$$

with

$$L_h(\boldsymbol{\beta}) = \prod_{j=1}^{N} \frac{\exp(\boldsymbol{\beta}^\top \mathbf{Z}_j)}{\sum_{l \in R_{hj}} \exp(\boldsymbol{\beta}^\top \mathbf{Z}_l)}$$

Here, the product in $L_h(\boldsymbol{\beta})$ is only taken over the event times from individuals in stratum h, and R_{hj} denotes the risk set at event time t_j in stratum h. If all relative risk parameters $\boldsymbol{\beta}$ are allowed to differ per strata, then the $L_h(\boldsymbol{\beta}) = L_h(\boldsymbol{\beta}_h)$ have nothing in common and fitting such a stratified Cox model boils down to fitting m different Cox models, i.e. one per stratum.

The results from a Cox model, which models effects of covariates on the hazard, can also be used to describe cumulative effects. For the moment, assume that only effects of time-fixed covariates have been modelled. If an individual has covariate values $\mathbf{Z}$, then, using (4), his or her survival curve is estimated as

$$\widehat{S}(t) = \exp\{-\widehat{\Lambda}_0(t)\, e^{\widehat{\boldsymbol{\beta}}^\top \mathbf{Z}}\} = \widehat{S}_0(t)^{\exp(\widehat{\boldsymbol{\beta}}^\top \mathbf{Z})} \tag{9}$$

with $\widehat{S}_0(t) = \exp(-\widehat{\Lambda}_0(t))$ the estimated baseline survival curve.

In the so-called counting process approach for right censored survival data, the number and type of events an individual experiences during his or her follow-up are counted. This approach allows for a nice representation of standard survival data, which is easily extended to more complex situations. With standard survival data, there is only one type of event and the number of events is either zero or one. An individual's survival data is expressed by three variables: the time the individual becomes at risk (`entry time`), the time the individual experiences the event or is censored (`event time`) and a variable denoting whether the event time is observed or censored (`status`). Consider the following example:

id	entry time	event time	status
1	0.0	4.3	1
2	0.0	5.6	0
3	3.4	7.7	1

The first individual experienced the event at time 4.3, and had been in follow-up since his time origin (e.g. transplant or HIV infection). The second individual was censored at time 5.6. If all individuals had been in follow-up from the time origin until the event or censoring, the `entry`

DOI: 10.1002/sim

time column would not be needed, since it would have the value zero for each individual. However, by including this extra time column, late entry (left truncation) with a known time origin can be described as well. The time value 3.4 of the third individual describes that the event determining his time origin occurred 3.4 time units before he came under follow-up. Moreover, time-dependent covariates are described in exactly the same way, with the inclusion of a column describing the value of the time-dependent covariate. For example, an individual that changes covariate value during follow-up, say at time 5.6, and experiences the event at time 7.7, is described as

```
id  start time  stop time  status  covar value
 4         0.0        5.6       0       A
 4         5.6        7.7       1       B
```

In the following sections, we will see some further extensions of this basic representation.

2.1. The independence assumption

Often, independence between the event and censoring distribution is assumed without further consideration, but may easily fail to be true. Reasons for the occurrence of right censored event times can be categorized as:

End of study: Since calendar time restricts observation to events that occurred in the past, an event time may be right censored because the individual has not been followed long enough yet. This is also called administrative censoring.
Loss to follow-up: The person has left the study, e.g. because of migration or study fatigue. He may have experienced the event already, but this information is missing.
Competing risk: Another event has occurred, which prevents occurrence of the event of interest.

If censoring is caused by end of study, we can in general safely assume that the censoring mechanism is independent of disease progression. In the other two situations (loss to follow-up and competing risks), one should be more cautious.

The censoring time due to loss to follow-up is negatively correlated with the event time when healthy participants feel less need for medical services offered from the study, and therefore quit. Censoring these individuals when they leave the study will cause a downward bias of the estimated survival curve, i.e. it will overestimate the probability to experience the event, since individuals with worse prognosis are assumed to be representative for the censored individuals. The censoring time is positively correlated with the event time when persons with advanced disease progression are more likely to leave the study. A reason may be that they have become too ill for further follow-up or that they return to their country of birth to spend the last period with their family. Here, censoring these individuals will cause an upward bias of the survival curve. Sometimes, extra information is available after drop-out, for example through registries. Using this information may decrease or remove bias, if selection of post-drop-out information is done in a proper way, which depends on the situation. Hoover *et al.* [11] considered censoring strategies with post-drop-out ascertainment and the resulting bias in parameter estimates in more detail. Although it is described in the context of HIV/AIDS cohort studies, the results apply more generally.

DOI: 10.1002/sim

3. COMPETING RISKS

3.1. Introduction

Competing risks concern the situation where more than one cause of failure is possible. If failures are different causes of death, only the first of these to occur is observed. In other situations, observations after the first failure may be observable, but not of interest. We can represent a competing risks model graphically with an initial state (alive or more generally event-free) and a number of different endpoints, as shown in Figure 1.

A number of examples from the medical field include:

1. One may have several endpoints which are of equal interest. For instance in bone marrow transplantation, death from different kinds of infections (bacterial, viral, fungal) are possible, as well as death due to relapse, graft-versus-host disease (GvHD) or other causes.
2. In cancer, death due to cancer may be of interest, and death due to other causes (surgical mortality, old age) are competing risks. Alternatively, one could be interested in time to relapse, where death due to any cause is a competing risk.
3. Interest is in the time from HIV infection to AIDS diagnosis (the incubation time) and whether this differs by risk group. Among injecting drug users, about 20 per cent of the HIV infected individuals dies before an AIDS diagnosis. Here, death before AIDS is a competing risk.
4. If one is interested in the time to staphylococcus infection during hospital stay in patients with burn wounds, censoring may occur due to death or hospital discharge. After hospital discharge, staphylococcus infection may still occur, but under completely different circumstances. One is then interested in the probability of infection during hospital stay. The competing event hospital discharge is non-fatal, but prevents the event of interest to occur *as a first event*.

Examples in other fields include failure of different components in a system in industrial reliability testing or time to part- or full-time employment in econometrics.

The subject of competing risks goes as far back as the 18th century, when Bernoulli [12] studied the possible consequences of eradication of smallpox on mortality rates. Indeed, the problem of estimation of failure probabilities after elimination (or modification) of one of the competing risks has been of great importance and has been the subject of much debate in the 1970s [13, 14].

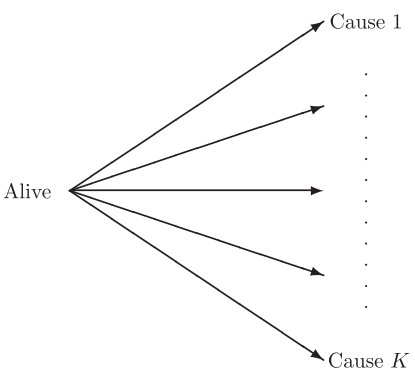

Figure 1. A competing risks situation with K causes of failure.

DOI: 10.1002/sim

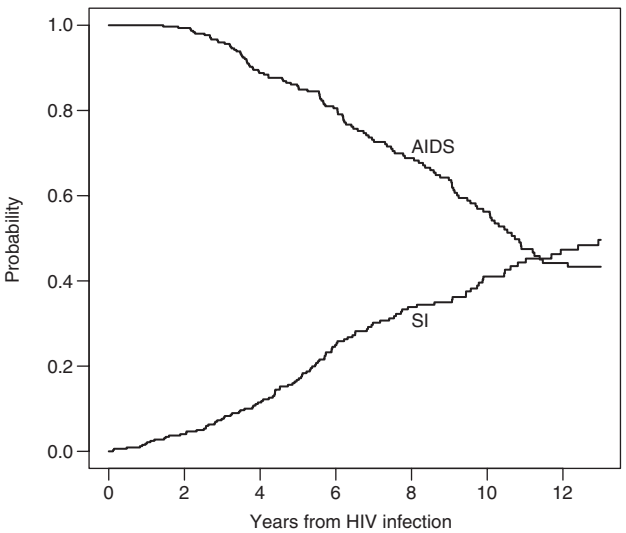

Figure 2. Estimated survival curve for AIDS and probability of SI appearance, based on the naive Kaplan–Meier estimator.

The central criticism is the assumption that upon removal of one cause of failure, the risks of failure of the remaining causes is unchanged. While this may be a reasonable assumption in the industrial setting, in human studies it will rarely be true.

For illustration of several concepts and techniques we will use data from 329 homosexual men from the Amsterdam Cohort Studies on HIV infection and AIDS [15]. During the course of HIV infection, the so-called syncytium inducing (SI) HIV phenotype appears in many individuals. Prognosis is strongly impaired after the appearance of this SI phenotype [16]. Little is known about factors that induce the appearance of SI phenotype. When analysing time to SI appearance before AIDS diagnosis, AIDS acts as a competing event.

In the first example above, each failure type is equally important. The other examples are more typical: one failure type can be singled out as the event of interest, while the remaining failure types are of less importance. One is then interested in the probability of failing from the cause of interest in the presence of competing risks (or, as in the first example, each of the death causes in turn is the cause of interest, with all the other death causes taken as competing risks). One method that is often used to estimate this failure probability is the Kaplan–Meier estimate, where the failures from the competing causes are treated as censored observations. This *naive Kaplan–Meier*, as we shall call it, is biased, however. Before discussing the reasons for this bias and ways to correctly estimate the failure probabilities, we first illustrate the bias by considering the data described above. For time to AIDS, all individuals in which SI phenotype appeared first were treated as censored, while for SI appearance, all AIDS diagnoses were treated as censored. Figure 2 shows the naive Kaplan–Meier estimates, where the Kaplan–Meier estimate of AIDS is represented as a survival curve, that of SI appearance as a probability distribution function (one minus survival). After 13 years of follow-up, the estimated probabilities of AIDS and SI appearance are 0.567 and 0.496, respectively. The curves of AIDS and SI appearance cross after 11 years, which means that the estimated probabilities of AIDS and SI appearance sum to more than one, which is clearly impossible, since in a competing risks context AIDS and SI appearance are disjoint first events.

DOI: 10.1002/sim

The basic issue in competing risks models that results in the bias of the naive Kaplan–Meier estimator is the violation of one of the assumptions underlying the Kaplan–Meier estimator: the assumption of independence of the censoring distribution, i.e. the distribution of the time to the competing events. If the competing event time distributions were independent of the distribution of time to the event of interest, this would imply that at each point in time the hazard of the event of interest is the same for subjects that have not yet failed and are still under follow-up as for subjects that have experienced a competing event by that time. However, a subject that is censored because of failure from a competing risk will with certainty NOT experience the event of interest. Since subjects that will never fail are treated as if they could fail (they are censored), the naive Kaplan–Meier overestimates the probability of failure (and hence underestimates the corresponding survival probability). The bias is greater when the competition is heavier, i.e. when the hazard of the competing events is larger. This is different from censoring due to end of study or loss to follow-up. In the latter situations, individuals may still fail at a later time point. One may argue that the naive Kaplan–Meier estimator describes what would happen if the competing event could be prevented to occur, creating an imaginary world in which an individual remains at risk for failure from the event of interest. This touches on the 1970s debate, since usually there is some biological mechanism that influences occurrence of both events, and changing the mechanism behind the competing event will also change the risk of the event of interest, i.e. time to the event of interest and time to the competing event are not independent. Hence this would be a completely different hypothetical situation about which we are not able to say anything. For an alternative explanation of the bias of the naive Kaplan–Meier estimator, see Reference [17].

3.2. Approaches to competing risks

The observable data in competing risks models is represented by the time of failure T, the cause of failure D, and possibly a covariate vector $\mathbf{Z}$, which we shall ignore for the moment. Inference therefore is to be based on the joint distribution of T and D, possibly given $\mathbf{Z}$. The fundamental concept in competing risks models is the *cause-specific hazard* function, the hazard of failing from a given cause in the presence of the competing events

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{\text{Prob}(t \leqslant T < t + \Delta t,\, D = k | T \geqslant t)}{\Delta t} \tag{10}$$

The cause-specific hazard is estimable from the data, see (17) below, and constitutes all relevant information that can be observed from the data. Also, anything that can be derived uniquely from the cause-specific hazard can be estimated.

Early approaches viewed competing risks models as a multivariate failure time model, where each individual is assumed to have a potential failure time for each type of failure. The earliest of these failures is actually observed and the others are latent. Let $\widetilde{T}_k$ denote the time to failure of cause k. We only observe $T = \min\{\widetilde{T}_k\}$ and D. Here D is an index variable, which specifies which event happened first. If some individuals are censored for all events by end of study or loss to follow-up, they have $D = 0$, and an extra censoring distribution $C \sim G$ is introduced, which is assumed to be independent of all the other events.

The latent failure time approach focused on the joint distribution of the times to the K different events, as described by the joint survival function

$$\overline{S}(t_1, \ldots, t_K) = \text{Prob}(\widetilde{T}_1 > t_1, \ldots, \widetilde{T}_K > t_K)$$

DOI: 10.1002/sim

The marginal distribution $S^k(t) = \text{Prob}(\widetilde{T}_k > t) = \overline{S}(0, \ldots, 0, t, 0, \ldots, 0)$ then defines a marginal hazard function as in (1). A fundamental problem with this approach is that, without additional assumptions, the joint survival function is not identifiable from the observed data (a single failure time for each subject). As already noted by Cox [18] and studied in detail by Tsiatis [19], for any joint survival function with arbitrary dependence between the different failure time distributions, one can find a different joint survival function with independent failure time distributions, *which has the same cause-specific hazards*. The implications of this are that the joint survival function is not identifiable, nor are the marginal distributions. It is even impossible to test for independence of the marginal failure time distributions! Sometimes extra information is available, for instance the value of some marker of progression was measured just before the competing event occurred. This marker may provide extra information on the dependence of the competing event. Now the problem is shifted to the impossibility of testing for independence conditionally on the value of the marker.

Anything that can be uniquely determined by the cause-specific hazards is estimable. Define the cumulative cause-specific hazard by

$$\Lambda_k(t) = \int_0^t \lambda_k(s)\,\mathrm{d}s$$

and define

$$S_k(t) = \exp(-\Lambda_k(t))$$

Note that, although $S_k(t)$ can be estimated, it should not be interpreted as a marginal survival function; it only has this interpretation if the competing event time distributions and the censoring distribution are independent. In that case, the marginal distribution describes the event time distribution in the situation that the competing events do not occur. Furthermore, define

$$S(t) = \exp\left(-\sum_{k=1}^{K} \Lambda_k(t)\right) \tag{11}$$

This survival function does have an interpretation; it is the probability of not having failed from any cause at time t. The *cumulative incidence function* of cause k, $I_k(t)$, is defined by the probability $\text{Prob}(T \leqslant t, D = k)$ of failing from cause k before time t. It can be expressed in terms of the cause-specific hazards as

$$I_k(t) = \int_0^t \lambda_k(s) S(s)\,\mathrm{d}s \tag{12}$$

Several alternative names have been used for this function, for example 'crude cumulative incidence function' or 'subdistribution function'. The latter name has its origin in the fact that the cumulative probability to fail from cause k remains below one, $I_k(\infty) = \text{Prob}(D = k)$, hence it is not a proper probability distribution.

Note that, as events from causes other than k are treated as censored, the naive Kaplan–Meier estimate of the probability of failing from cause k before or at time t is estimating

$$1 - S_k(t) = \int_0^t \lambda_k(s) S_k(s)\,\mathrm{d}s$$

DOI: 10.1002/sim

The difference with the cumulative incidence function $I_k(t)$ from equation (12) is that $S(s)$ is replaced by $S_k(s)$. Since $S(t) \leqslant S_k(t)$, we have $I_k(t) \leqslant 1 - S_k(t)$, with equality at t if there is no competition, i.e. if $\sum_{j=1, j\neq k}^{K} \Lambda_j(t) = 0$, again showing the bias in the naive Kaplan–Meier estimator.

The cumulative incidence function is also used extensively in calculating state and prediction probabilities in multi-state models. In fact, as we shall see in the next section, competing risks models are a special case of multi-state models and the cumulative incidence approach has been termed the multi-state approach to competing risks [1].

We now turn to estimation of the cumulative incidence functions. Let $0 < t_1 < t_2 < \cdots < t_N$ be the ordered distinct time points at which failures *of any cause* occur. Let d_{kj} denote the number of patients failing from cause k at t_j, and let $d_j = \sum_{k=1}^{K} d_{kj}$ denote the total number of failures (from any cause) at t_j. In the absence of ties only one of the d_{kj} equals 1 for a given j, and $d_j = 1$. The formulas are also valid, however, in the presence of ties. As in Section 2, let n_j be the number of patients at risk (i.e. that are still in follow-up and have not failed *from any cause*) at time t_j. The overall survival probability $S(t)$ at t can be estimated, without considering the cause of failure, by the Kaplan–Meier estimator

$$\widehat{S}(t) = \prod_{j:t_j \leqslant t} \left(1 - \frac{d_j}{n_j}\right) \tag{13}$$

familiar from equation (7). As in Section 2, consider a discretized version of the cause-specific hazard of equation (10),

$$\lambda_k(t_j) = \text{Prob}(T = t_j, D = k | T > t_{j-1}) \tag{14}$$

Similar to (5), this quantity would be estimated by

$$\widehat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j}$$

the proportion of subjects at risk that fail from cause k. Note that (13) can also be written down as

$$\widehat{S}(t) = \prod_{j:t_j \leqslant t} \left(1 - \sum_{k=1}^{K} \widehat{\lambda}_k(t_j)\right) \tag{15}$$

The unconditional probability of failing from cause k at t_j, $p_k(t_j) = \text{Prob}(T = t_j, D = k)$ is the product of the hazard and the probability of being event-free at t_j, and is estimated as

$$\widehat{p}_k(t_j) = \widehat{\lambda}_k(t_j)\widehat{S}(t_{j-1}) \tag{16}$$

Finally, the cumulative incidence $I_k(t)$ of cause k at t is estimated as the sum of these terms for all time points before t; in summary

$$\widehat{I}_k(t) = \sum_{j:t_j \leqslant t} \widehat{p}_k(t_j), \quad \widehat{p}_k(t_j) = \widehat{\lambda}_k(t_j)\widehat{S}(t_{j-1}), \quad \widehat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j} \tag{17}$$

Table I illustrates the steps in estimating the cumulative incidence functions for AIDS and SI appearance in the SI data. For example, at time $t_j = 0.112$, SI appeared in one individual. The estimated overall survival at the previous time point is 1 (there was no earlier event), and the

DOI: 10.1002/sim

Table I. Illustration of the steps used in estimating the cumulative incidence functions for AIDS and SI appearance in the SI data.

				Cause 1 (AIDS)				Cause 2 (SI appearance)			
Time	No. at risk	Total no. of failures	Estimated overall survival	No. of failures	Estimated failure rate	Estimated failure probability	Estimated cumulative incidence	No. of failures	Estimated failure rate	Estimated failure probability	Estimated cumulative incidence
t_j	n_j	d_j	$\widehat{S}(t_j)$	d_{1j}	$\widehat{\lambda}_1(t_j)$	$\widehat{p}_1(t_j)$	$\widehat{I}_1(t_j)$	d_{2j}	$\widehat{\lambda}_2(t_j)$	$\widehat{p}_2(t_j)$	$\widehat{I}_2(t_j)$
0.112	329	1	0.9970	0	0	0	0	1	0.0030	0.0030	0.0030
0.137	328	1	0.9939	0	0	0	0	1	0.0030	0.0030	0.0061
0.142	327	0	0.9939	0	0	0	0	0	0	0	0.0061
0.148	326	0	0.9939	0	0	0	0	0	0	0	0.0061
0.474	325	1	0.9909	0	0	0	0	1	0.0031	0.0031	0.0091
⋮			⋮	⋮			⋮	⋮			⋮
1.437	310	0	0.9723	0	0	0	0	0	0	0	0.0277
1.440	309	1	0.9691	1	0.0032	0.0031	0.0031	0	0	0	0.0277
1.457	308	0	0.9691	0	0	0	0.0031	0	0	0	0.0277
1.462	307	1	0.9660	0	0	0	0.0031	1	0.0033	0.0032	0.0309
1.503	306	1	0.9628	0	0	0	0.0031	1	0.0033	0.0032	0.0340

Note that, due to rounding errors, the rounded numbers do not always exactly add up.

DOI: 10.1002/sim

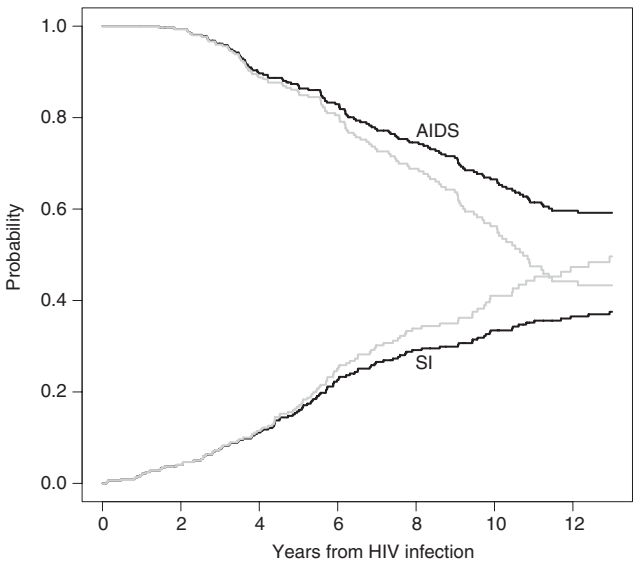

Figure 3. Estimates of probabilities of AIDS and SI appearance, based on the naive Kaplan–Meier (grey) and on cumulative incidence functions (black).

estimate of the failure rate $\hat{\lambda}_2(0.112)$ is $\frac{1}{329} = 0.0030$. Since the overall survival is one, 0.0030 is also the estimate of the unconditional probability $\hat{p}_2(0.112)$. The first AIDS event occurs at time 1.440. At this time, 309 patients are at risk. The estimated overall survival at the previous time point 1.437 is 0.9723, and the estimate of the failure rate $\hat{\lambda}_1(1.440)$ is $\frac{1}{309} = 0.0032$, yielding $0.9723 \times 0.0032 = 0.0031$ for the estimated unconditional failure probability.

Figure 3 shows the estimates of the probabilities of AIDS and SI appearance for all patients in the SI data, using the same representation as Figure 2. In grey are the estimates based on the naive Kaplan–Meier, in black those based on the cumulative incidence functions. Recall that the estimates based on Kaplan–Meier after 13 years of follow-up are 0.567 and 0.496, cumulative incidence estimates are 0.408 and 0.375, for AIDS and SI appearance, respectively. Figure 4 shows the estimated cumulative incidence curves again, laid out in a different way. They are stacked; the bottom curve shows $\widehat{I}_1(t)$, the top curve $\widehat{I}_1(t) + \widehat{I}_2(t)$, where $\widehat{I}_1(t)$ and $\widehat{I}_2(t)$ are the estimates of the cumulative incidence functions of AIDS and SI appearance respectively. The distances between adjacent curves now correspond to the probabilities of the events. This representation is particularly useful for displaying more than two competing risks and for multi-state models.

If there are only competing events and no censoring or left truncation, then the estimate of the cumulative incidence function reduces to a very simple form. At time t, the estimate divides the cumulative number of events of type k until time t by the total sample size. Hence, individuals remain in the denominator, even though they have experienced a competing event.

3.3. Modelling and estimating covariate effects

Just like in standard survival analysis, the effect of one or two binary covariates is most easily investigated by estimating cumulative incidence curves non-parametrically and testing whether the

DOI: 10.1002/sim

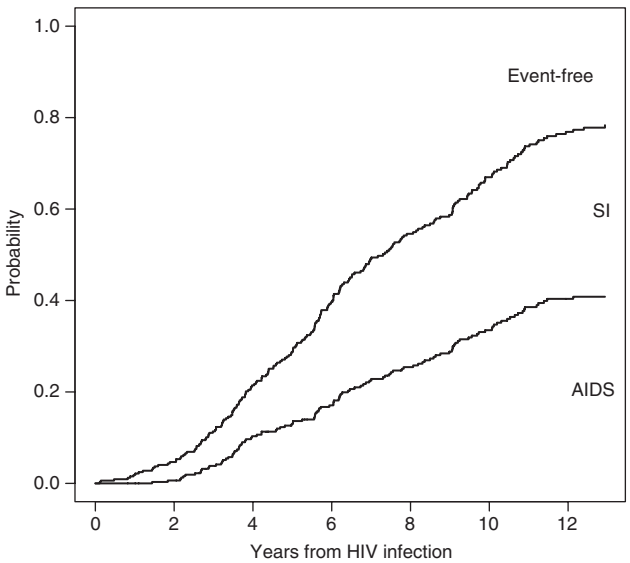

Figure 4. Cumulative incidence curves of AIDS and SI appearance. The cumulative incidence functions are stacked; the distance between two curves represent the probabilities of the different events.

curves differ by covariate value. Gray [20] developed a log-rank type test for equality of cumulative incidence curves.

In this subsection we shall illustrate the use of R [10] in carrying out some of the regression analyzes based on the SI data set. A specific deletion in the C–C chemokine receptor 5 gene (CCR5 Δ32) has been associated with reduced susceptibility to HIV infection and delayed AIDS progression. Since NSI viruses use CCR5 for cell entry, whereas SI viruses can also use C-X-C chemokine receptor 4 (CXCR4), the latter virus type may have an advantage in persons with the deletion. Therefore, we investigate whether in persons with the deletion the SI phenotype appears more rapidly. This question has been addressed using standard survival analysis techniques [21], which implicitly assumed that a switch to SI and progression to AIDS are independent mechanisms. The CCR5 genotype is incorporated in the SI data set through the covariate `ccr5`. Persons without the deletion ('wild type') have `WW`, the reference category, whereas individuals who have the deletion on one of the chromosomes have `WM` (individuals with the deletion on both chromosomes were not present in our data).

As a preliminary, we introduce two ways of representing the same data. The first of these is the standard way of representing competing risks data. Consider the first four patients of the SI data set, in regular format:

```
  patnr   time status      cause ccr5
1     1  9.106      1       AIDS   WW
2     2 11.039      0 event-free   WM
3     3  2.234      1       AIDS   WW
4     4  9.878      2         SI   WM
```


DOI: 10.1002/sim

Here a single `time` and `cause` variable are used to indicate time of failure (or censoring) and cause of failure. The variable `status` is just a numeric representation of `cause`. The whole data set represented in this format will be called `si`. An alternative way of representing the same data is in *long* format (the SI data set in long format is called `silong`). We will see later that this representation allows for more flexibility in modelling the effect of covariates. The same data in long format look like this:

	patnr	time	status	stratum	cause	ccr5	ccr5.1	ccr5.2
1	1	9.106	1	1	AIDS	WW	0	0
2	1	9.106	0	2	SI	WW	0	0
3	2	11.039	0	1	AIDS	WM	1	0
4	2	11.039	0	2	SI	WM	0	1
5	3	2.234	1	1	AIDS	WW	0	0
6	3	2.234	0	2	SI	WW	0	0
7	4	9.878	0	1	AIDS	WM	1	0
8	4	9.878	1	2	SI	WM	0	1

If there are K competing events, each individual needs K rows in the new data file, one for each possible cause of failure. A column (`cause` in the example) is used to denote the event type or failure cause that the row refers to. The value of the `time` variable is identical over the K rows of an individual. The `status` variable changes. Instead of values $0, 1, \ldots, K$, it now has the value 1 if the corresponding event type is the one that occurred, and it has the value 0 otherwise. Any covariates are simply replicated for each patient over the K rows of that individual. We have also introduced two extra dummy variables `ccr5.1` and `ccr5.2`. They have the value 0 except for mutant (`WM`) genotypes for the `cause` that they correspond to (i.e. for a patient with the mutant genotype, `ccr5.1` $= 1$ for the first cause, 'AIDS', `ccr5.2` $= 1$ for the second cause, 'SI'). They are what Andersen *et al.* [22] call type-specific covariates.

3.3.1. Regression on cause-specific hazards. If the covariate is continuous or the simultaneous effect of several covariates on cause-specific failure is of interest, a competing risks analogue of a Cox proportional hazards model seems the most logical choice [23]. Since the cause-specific hazards are identifiable, regression on the cause-specific hazards is possible. In proportional hazards regression on the cause-specific hazards, we model the cause-specific hazard of cause k for a subject with covariate vector $\mathbf{Z}$ as

$$\lambda_k(t|\mathbf{Z}) = \lambda_{k,0}(t)\exp(\boldsymbol{\beta}_k^\top \mathbf{Z}) \tag{18}$$

where $\lambda_{k,0}(t)$ is the baseline cause-specific hazard of cause k, and the vector $\boldsymbol{\beta}_k$ represents the covariate effects on cause k. The analysis is completely standard, but the interpretation requires caution, as we shall see later. At each time some person moves to state k, the covariate values of this individual are compared with the covariates of all other individuals still event-free and in follow-up. Persons who move to another state are censored at their transition time.

As an example, let us look at the effect of CCR5 (classified as wild-type (`WW`) or mutant (`WM`)) on AIDS and SI appearance, using (18). A total of 259 out of 324 patients (80 per cent) had the wild-type variant, while 65 patients (20 per cent) had the mutant variant. Five patients had unknown

DOI: 10.1002/sim

CCR5-genotype.

```
> coxph(Surv(time, status == 1) ~ ccr5, data = si)
Call: coxph(formula = Surv(time, status == 1) ~ ccr5, data = si)

          coef exp  (coef)  se(coef)      z        p
  ccr5WM  -1.24      0.291     0.307  -4.02  5.7e-05

  Likelihood ratio test=22 on 1 df, p=2.76e-06 n= 324

> coxph(Surv(time, status == 2) ~ ccr5, data = si)
Call: coxph(formula = Surv(time, status == 2) ~ ccr5, data = si)

          coef exp(coef) se(coef)    z  p
  ccr5WM -0.254    0.776   0.238 -1.07 0.29

  Likelihood ratio test=1.19 on 1 df, p=0.275 n= 324
```

Some familiarity with R, in particular with the use of formulas, and with the `survival` library by Therneau [24] is needed to fully understand the code. For this, we refer to one of the tutorials on the Comprehensive R Archive Network (`http://cran.r-project.org/`). However, what these statements do is fit a Cox proportional hazards model with `ccr5` as sole covariate, first using `status` = 1 (AIDS) as event (so censoring SI appearances), then using status = 2 (SI appearance) as event (censoring AIDS events). The estimated coefficient for the mutant with respect to the wild-type variant for AIDS was -1.24 (SE 0.31), giving a significant protective effect of the mutant variant (hazard ratio (HR) = 0.29, $P<0.0001$). The effect of CCR5 on SI appearance was not significant (coefficient -0.25, SE 0.24, HR 0.78, $P=0.29$).

The same model as before, with different effects of CCR5 on AIDS and SI appearance, can also be fitted using data in long format. In fact, this can be done in a number of ways. One is to use only subsets of the data corresponding to the cause of failure of interest:

```
> coxph(Surv(time, status) ~ ccr5, data = silong, subset=cause=="AIDS")
```

and

```
> coxph(Surv(time, status) ~ ccr5, data = silong, subset=cause=="SI")
```

Another is to use the dummies `ccr5.1` and `ccr5.2`, to obtain an attractively simple analysis:

```
> coxph(Surv(time, status) ~ ccr5.1 + ccr5.2 + strata(cause),
  data = silong)
Call:
coxph(formula = Surv(time, status) ~ ccr5.1 + ccr5.2 + strata(cause),
  data = silong)

              coef exp(coef) se(coef)     z       p
  ccr5.1WM -1.236     0.291    0.307 -4.02 5.7e-05
  ccr5.2WM -0.254     0.776    0.238 -1.07 2.9e-01

  Likelihood ratio test=23.2 on 2 df, p=9.3e-06 n=648
```


DOI: 10.1002/sim

The $n = 648$ mentioned here equals the number of rows (two times 324) in the long data set without missing data (a warning from R that 10 observations were not used because of missing covariates has been removed from the output). The same model can also be fitted by adding an interaction term between the cause stratum variable and age.

```
> coxph(Surv(time, status) ~ ccr5 * cause + strata(cause),
  data = silong)
Call:
coxph(formula = Surv(time, status) ~ ccr5 * cause + strata(cause),
   data = silong)

                   coef exp(coef) se(coef)     z       p
  ccr5WM         -1.236     0.291    0.307 -4.02 5.7e-05
  causeSI           NA        NA    0.000    NA      NA
  ccr5WM:causeSI  0.982     2.669    0.389  2.53 1.2e-02

  Likelihood ratio test=23.2 on 2 df, p=9.3e-06 n=648
```

Now we see the advantage of the use of the long format. The notation in (18) allows the effect of the covariates to be different for each failure cause. Use of the long format makes it possible to assume that the effects of CCR5 are identical for the different causes and to test for equality of the effects of CCR5 on AIDS and SI appearance. The coefficient -1.236 is (as before) for the effect of CCR5 on AIDS. The deviant coefficient 0.982 now represents the *difference* in the effect of CCR5 on the two cause-specific hazards. The CCR5 genotype by cause interaction term is significant, indicating that the effect of CCR5 is quite different on AIDS and SI appearance. The effect of CCR5 on SI appearance is thus given by $-1.236 + 0.982 = -0.254$, as before. Note that the second row with NA's in the output above is caused by the fact that the cause main effect cannot be estimated, since the baseline cause-specific hazards are both freely estimated.

Although not applicable here, if we were to assume that the effect of CCR5 on the two cause-specific hazards is equal, we could use

```
> coxph(Surv(time, status) ~ ccr5 + strata(cause), data = silong)
```

There are two alternative ways yielding the same result. First, it can be shown, by carefully writing out the partial likelihood, that the strata can be left out.

```
> coxph(Surv(time, status) ~ ccr5, data = silong)
```

The reason is that in both strata the risk sets as well as the covariate values (here ccr5) are equal. Second, since the strata term is not needed, we can use si in original format:

```
> coxph(Surv(time, status != 0) ~ ccr5, data = si)
```

Finally, we show the analyzes under the assumption that the baseline cause-specific hazards are proportional. Now cause is not used as stratum, but as another covariate for which a relative risk parameter is estimated. The R code for this is given by

```
> coxph(Surv(time, status) ~ ccr5.1 + ccr5.2 + cause, data = silong)
Call:
coxph(formula = Surv(time, status) ~ ccr5.1 + ccr5.2 + cause,
   data = silong)
```


DOI: 10.1002/sim

```
          coef exp(coef) se(coef)     z       p
ccr5.1 -1.166     0.311    0.306 -3.81 0.00014
ccr5.2 -0.332     0.718    0.237 -1.40 0.16000
causeSI -0.184    0.832    0.148 -1.25 0.21000

Likelihood ratio test=21.5 on 3 df, p=8.12e-05 n=648
```

The coefficient -0.184 and its hazard ratio 0.832 would indicate that (under the assumption of the cause-specific hazards being proportional) the baseline cause-specific hazard of SI appearance is somewhat smaller than that of AIDS, though not significant ($P = 0.21$). Even though the assumption of proportional baseline cause-specific hazards will often be unrealistic, this proportional risk model has the nice property that the probability of an individual failing of cause k follows a logistic model [23].

The covariate effects in (18) are proportional for the cause-specific hazards. In the absence of competing risks this would mean that the survival functions for different values of the covariates were related through a simple formula. If S_1 and S_2 are the survival functions for covariate values $\mathbf{Z}_1$ and $\mathbf{Z}_2$, then (cf. also (9))

$$S_2(t) = S_1(t)^{\exp(\boldsymbol{\beta}^\top(\mathbf{Z}_2 - \mathbf{Z}_1))} \tag{19}$$

However, in the presence of competing risks, when the effect of the same covariates are also modelled for other causes of failure, this relation does not extend to cumulative incidence functions. The reason is that the cumulative incidence function for cause k not only depends on the hazard of cause k, but also on the hazards of all other causes (recall the definition of the cumulative incidence function from (12)). Hence the relation of the cumulative incidence functions of cause k for two different covariate values not only depends on the effect of the covariate on cause k, but also on the effects of the covariate on all other causes and on the baseline hazards of all other causes. As a result, the simple effect of a covariate on the cause-specific hazard of cause k can be quite unpredictable when expressed in terms of the cumulative incidence function.

Figure 5 shows the estimated cumulative incidence functions for both wild-type and mutant variants of CCR5 based on the above regression model and formulas (15) and (18), for AIDS (left) and for SI appearance (right). While the protective effect of the mutant WM on AIDS is clear, on close inspection it is apparent that the effect of CCR5 on the probability of SI appearance is not quite as expected from a standard situation without competing risks. In the latter situation, since the hazard ratio is 0.78, the patients with the mutant genotype would have a consistently lower probability of SI appearance, and the difference in SI probabilities between mutant and wild-type would increase with time. Here, although initially the probability of SI appearance is indeed lower for the mutant WM, after approximately 9 years the difference decreases rather than increases, and after 11 years the cumulative incidence functions of AIDS and SI appearance cross. This is caused by the fact that although the hazard of SI appearance is lower for WM, the hazard of AIDS is also lower for WM, and the effect is much stronger for AIDS. Both the effect of the covariate on the competing risk and the baseline hazard of the competing risk influence the effect of the covariate on the cumulative incidence of the event of interest. The fact that the baseline hazard of the competing risk matters is perhaps unexpected, so we illustrate the fact that the baseline hazard of AIDS (i.e. corresponding to the wild-type WW) plays an important role here in two ways.

In Figure 6, we have considered a somewhat idealized situation, where we have a population of 10 000 individuals with the wildtype WW and 10 000 individuals with the mutant WM genotype.

DOI: 10.1002/sim

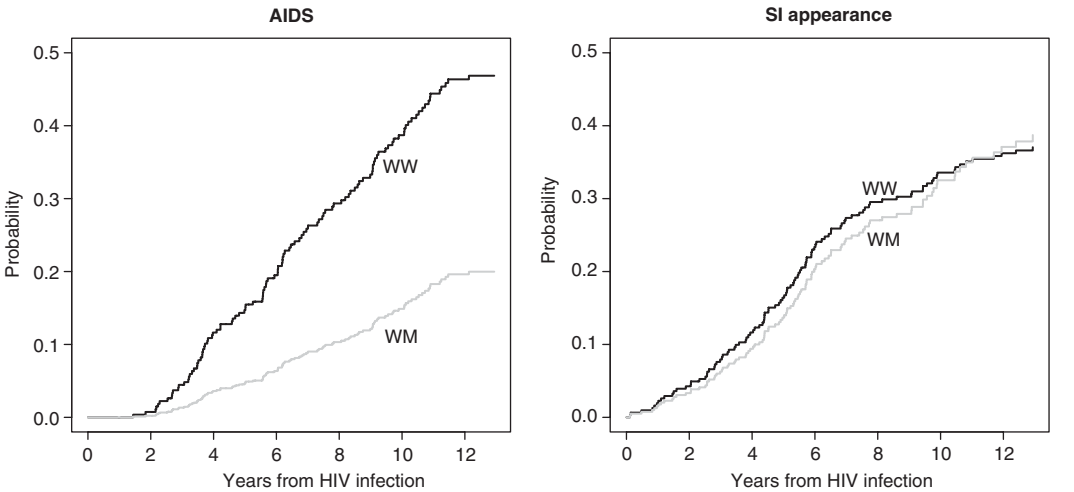

Figure 5. Cumulative incidence functions for AIDS (left) and SI appearance (right), for wild-type (WW) and mutant (WM) CCR5 genotype, based on a proportional hazards model on the cause-specific hazards.

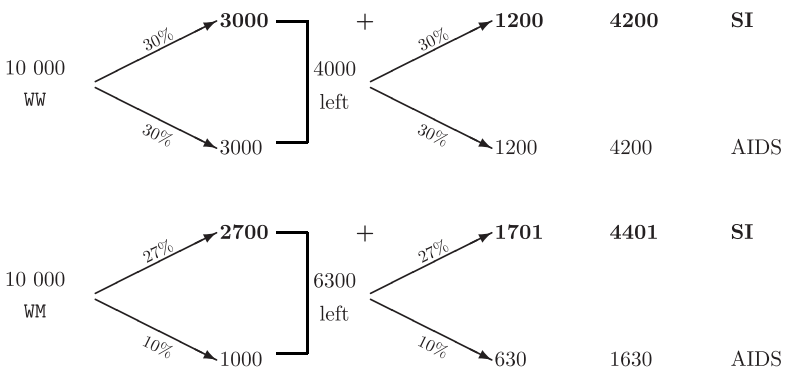

Figure 6. The difference between covariate effects on cause-specific hazards and cumulative incidence explained.

We assume that WW individuals have a constant failure rate of 30 per cent at discrete time points, for both endpoints. The mutation WM is protective for the cause-specific hazard to SI appearance (hazard ratio 0.90). However, it is even more protective for AIDS diagnosis (hazard ratio 0.33). This latter aspect causes more individuals to remain at risk after the first round for WM. Hence, in the second round, SI appears in more individuals with WM than in individuals with WW (1701 to 1200). As a result, after the second round, the cumulative incidence for SI appearance is higher for individuals with WM than for individuals with WW genotype. The second illustration of this phenomenon is through Figure 7, which shows what would happen if we were to change the baseline hazard of AIDS by multiplying the estimate from the data with different multiplication factors, while keeping everything else (the baseline cause-specific hazard of SI appearance,

DOI: 10.1002/sim

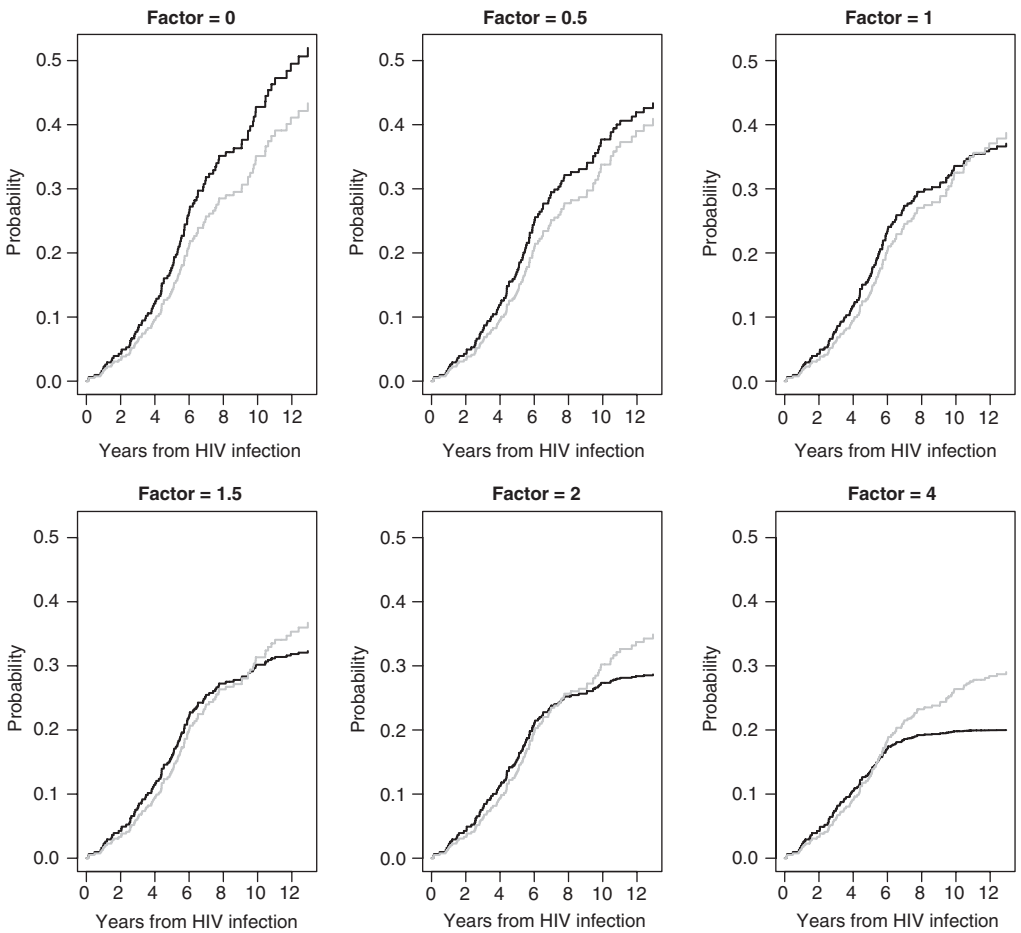

Figure 7. Cumulative incidence functions for SI appearance, for CCR5 wild-type WW (black) and mutant WM (grey). The baseline hazard of AIDS was multiplied with different factors, while keeping everything else the same.

and the effects of CCR5 on both cause-specific hazards) the same. The sub-plot with factor $= 0$ corresponds to the standard Cox regression in the absence of the competing risk 'AIDS'. Here the difference in probabilities of SI appearance between wild-type and mutant indeed increases with time. As the competition from AIDS is increased, the higher cause-specific hazard for SI appearance, $\lambda_{\mathrm{SI}}(s)$, for WW compared to WM is offset against an increasingly smaller contribution from the overall survival $S(s) = \exp(-(\Lambda_{\mathrm{AIDS}}(s) + \Lambda_{\mathrm{SI}}(s)))$ for WW, where the contribution of AIDS, $\Lambda_{\mathrm{AIDS}}(s)$, increases as the multiplication factor increases. At first this results in a crossing of the cumulative incidence curves (see e.g. factor $= 1$, this is not possible in the absence of competing risks), which occurs earlier with increasing multiplication factor. With factor $= 4$, the effect of CCR5 on the cumulative incidence of SI appearance is inverse to what the hazard ratio of 0.78 of WM with respect to WW seems to suggest.

DOI: 10.1002/sim

The use of long format, in particular in combination with the use of cause-specific dummies (`ccr5.1` and `ccr5.2` in our example) and stratified Cox regression offers great flexibility in modelling the effect of covariates on the cause-specific intensity rates, while using standard statistical software [25]. Several authors have suggested that robust estimates of standard errors should be used in order to correct for the correlation caused by multiplication of the data set (see e.g. [25]). However, each individual still has at most one event, so that standard estimates of the standard error do suffice (see also the discussion in Reference [7] and our online material).

If the number of competing events becomes large or if one of the events is rare, equality of effects or proportionality of baseline hazards may become a necessary assumption to prevent overfitting. The reduced rank proportional hazards model for competing risks, introduced in Fiocco *et al.* [26] may be helpful in such situations. For the special case of rank one such a reduced rank model is a proportional hazards model where each covariate has the same effect on all transitions except for proportionality coefficients. More generally, the reduced rank proportional hazards model of rank R requires the matrix of regression coefficient vectors, stacked horizontally column by column for different causes, to be of reduced rank R, smaller than the number of failure causes, K, and the number of covariates, p. It has the advantage of modelling each transition in a different way with fewer parameters, deals with transitions with rare events and overcomes the problem of over-fitting. Two applications of this method, to leukaemia-free patients surviving a bone marrow transplant [26] and to data from a breast cancer trial [27], led to interpretable results that made clear clinical sense but were not immediate from the full rank models.

3.3.2. Regression on cumulative incidence functions. In order to avoid the highly nonlinear effects of covariates on the cumulative incidence functions when modelling is done on the cause-specific hazards, Fine and Gray [28] introduced a way to regress directly on cumulative incidence functions. In analogy with the relation (2) between hazard and survival, they defined a *subdistribution hazard*

$$\overline{\lambda}_k(t) = -\frac{\mathrm{d}\log(1 - I_k(t))}{\mathrm{d}t} \tag{20}$$

This is not the cause-specific hazard. In terms of estimates of this quantity, the difference is in the risk set. For the cause-specific hazard, the risk set decreases at each time point at which there is a failure of another cause. For $\overline{\lambda}_k(t)$, persons who fail from another cause *remain in the risk set*. If there is no censoring, they remain in the risk set forever and once these individuals are given a censoring time that is larger than all event times, the analysis becomes completely standard. If there is censoring, they remain in the risk set until their potential censoring time, which is not observed if they experienced another event before. With administrative censoring, the potential censoring time is still known. If individuals may also be lost to follow-up, a censoring distribution is estimated from the data. Fine and Gray imposed a proportional hazards assumption on the subdistribution hazards:

$$\overline{\lambda}_k(t|\mathbf{Z}) = \overline{\lambda}_{k,0}(t)\exp(\boldsymbol{\beta}_k^\top \mathbf{Z}) \tag{21}$$

Estimation follows the partial likelihood approach used in a standard Cox model. In a later paper, Fine extended this idea to other link functions using an estimating equations approach. Using the R library `cmprsk` we obtain the following results (after removing the five subjects with missing

DOI: 10.1002/sim

CCR5 covariate values and making `ccr5` numeric).

```
> library(cmprsk)
> crr(si$time,si$status,si$ccr5) # for failures of type 1 (AIDS)
convergence: TRUE
coefficients:
[1] -1.004
standard errors:
[1] 0.295
two-sided p-values:
[1] 0.00066
> crr(si$time,si$status,si$ccr5,failcode=2) # for failures of type
  2 (SI)
convergence: TRUE
coefficients:
[1] 0.02359
standard errors:
[1] 0.2266
two-sided p-values:
[1] 0.92
```

The protective effect of the mutant `WM` genotype on AIDS is again apparent ($P = 0.0007$). Note that the effect of the mutant `WM` genotype on SI appearance has reversed compared to regression on cause-specific hazards, though it is very far from significant.

Figure 8 shows the predicted cumulative incidence curves for time to AIDS and time to SI appearance based on the Fine and Gray results. Note that the cumulative incidence curves of SI appearance for CCR5 wild-type and mutant do not cross and that the cumulative incidence curve

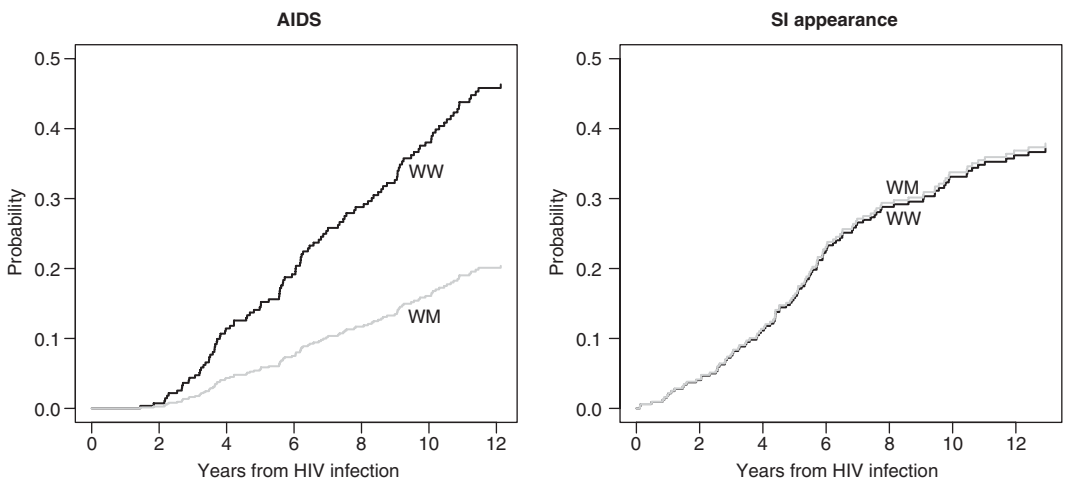

Figure 8. Cumulative incidence functions for AIDS (left) and SI appearance (right), for CCR5 wild-type (`WW`) and mutant (`WM`), based on the Fine and Gray model.

DOI: 10.1002/sim

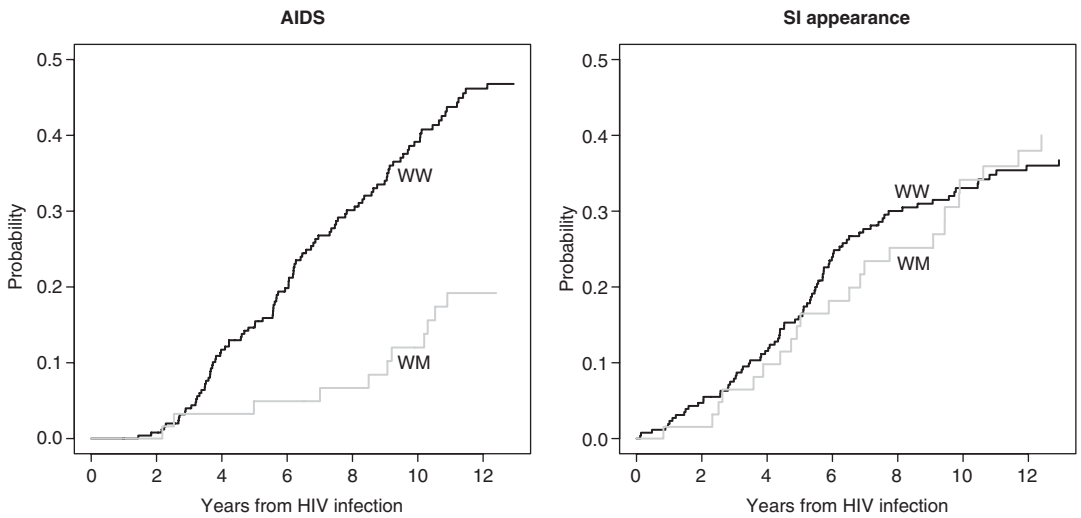

Figure 9. Non-parametric cumulative incidence functions for AIDS (left) and SI appearance (right), for CCR5 wild-type (`WW`) and mutant (`WM`).

of the mutant is above that of the wild-type. As far as we know the Fine and Gray regression does not yet allow the flexibility (e.g. in testing for or assuming equality of covariate effects across different causes) of regression on cause-specific hazards. Also, it is not clear how left truncated data or time-dependent covariates can be included in their approach.

We have presented the Fine and Gray method here as a way of repairing problems with proportional hazards regression on cause-specific hazards. We would like to stress that there is nothing fundamentally wrong with regression on cause-specific hazards. The problems lie in the fact that we are used to interpreting hazard ratios in the standard proportional hazards regression with a single endpoint as implying a qualitatively similar cumulative effect via a relation like (19). One should be aware that this relation is no longer true in the presence of competing risks; it does not mean that the model itself is incorrect. A straightforward way of judging the goodness-of-fit of the two approaches is by comparing the predicted cumulative incidence curves of the regression models with the non-parametric cumulative incidence curves obtained by applying (17) to the subset of CCR5 wild-type and mutant separately. Figure 9 shows these model-free cumulative incidence curves. Judging from Figure 9, particularly for SI appearance, the cumulative incidence curves of the proportional hazards regression model on cause-specific hazards (Figure 5) follow the non-parametric cumulative incidence curves quite closely, more so than the cumulative incidence curves from the Fine and Gray regression (Figure 8).

3.4. Software

Regression on cause-specific hazards can be performed in any package that includes the Cox proportional hazards model. An option to fit stratified Cox models needs to be included if we want to fit or test for equality of covariate effects for different transitions.

Cumulative incidence curves in a competing risks setting can be estimated in S-PLUS/R (`cmprsk` library), Stata (`stcompet.ado` module) and NCSS. The Stata web site also

DOI: 10.1002/sim

provides further explanations on fitting competing risks models (see `http://www.stata.com/support/faqs/stat/stmfail.html`). Rosthøj *et al.* [29] have written a set of SAS macros that allows to translate results from a Cox model on cause-specific hazards into cumulative incidence curves for some choice of covariate values (see `http://www.pubhealth.ku.dk/~pka`). It also calculates standard errors. A Cox null model without covariates can be used to obtain a single cumulative incidence curve for the whole group. The `R` package `mstate`, further mentioned in Section 4.6, can also be used for competing risks, and also implements the reduced rank approach of Fiocco *et al.* [26].

3.5. Summary and concluding remarks

We have seen that modelling the effect of covariates on cause-specific hazards may lead to different conclusions than modelling their effect on subdistribution hazards and cumulative incidence functions. The standard Cox model can be used to model the effect of covariates on the cause-specific hazards of the different endpoints. The data format used is basically the same as in a standard survival analysis with one endpoint (the long format is just a clever way of combining data for the different endpoints such that all can be analyzed at once). If we start with a Cox model on cause-specific hazards, we have the advantage of a wealth of theory that has been developed and software that has been written for this purpose. Cause-specific hazards as obtained from a Cox model can be translated into cumulative incidence curves through formula (17). The problem is that proportionality is lost and hence covariate effects on cumulative incidence curves can no longer be expressed by a simple number. The main lesson to be learned here is that to determine the effect of a covariate on the cumulative incidence of an event of interest it is also important to consider the competing risk(s) (both baseline and effect of covariate). Still, results from the Cox model do not provide an answer to the question what the effect of some covariate would have been on the cause-specific hazard if the competing risks were absent, unless the competing events are independent. Di Serio [30], in a simulation study, has shown that the estimated effect of a covariate may even be reversed if the dependence between two endpoints is caused by a common factor that is not included in the model, but is correlated with the covariate of interest. Regression on cumulative incidence curves allows to describe the effect of covariates through simple numbers. To our knowledge, software to fit these models has only been written in S-PLUS/R.

We have presented the most common approaches to the analysis of competing risks. Another approach to regression with competing risks is to use pseudo-observations, as explained in References [31, 32]. Sometimes, different endpoints occur at the same time. For example, relapse may occur at several locations simultaneously, or an HIV infected person may be diagnosed with several AIDS defining illnesses. See Reference [33] for some approaches to analyze such data. Sometimes, two different groups of endpoints can occur simultaneously, e.g. when two different classification schemes are used. Such data lead to the concept of multivariate competing risks [34].

4. MULTI-STATE MODELS

4.1. Introduction

The class of multi-state models forms an extension to that of competing risks models. Competing risks models deal with one initial state and several mutually exclusive absorbing states. Typically,

DOI: 10.1002/sim

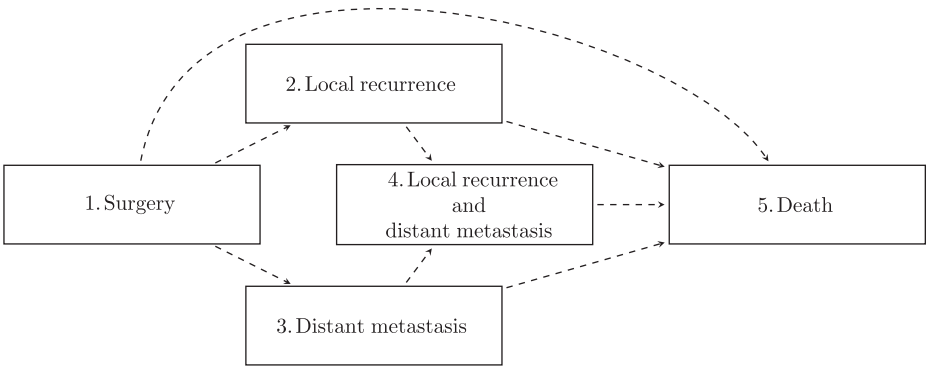

Figure 10. A multi-state model for breast cancer.

the disease or recovery process of a patient will also consist of intermediate events that can neither be classified as initial states nor as final states. This type of models is called multi-state models.

Before introducing the necessary terminology, let us consider a number of examples:

1. In many cancer studies, after surgery of the primary tumour, the tumour may recur in the vicinity of the primary tumour (local recurrence), or at distant locations (distant metastasis). These events may occur in any order (although local recurrence usually precedes distant metastasis) and patients may die before or after experiencing local recurrence or distant metastasis. Figure 10 illustrates a multi-state model that has been used to describe the disease process in a breast cancer study [35].
2. After bone marrow transplantation, patients may acquire acute graft-versus-host disease (GvHD), a reaction of the immune system in the donor graft against normal host tissues. The acute GvHD may become chronic. Patients may also relapse, either before or after acute GvHD, or die. Another intermediate event is the recovery of platelet count to normal levels. In the literature, many papers have appeared dealing with multi-state models on bone marrow transplantation, see e.g. References [36–39].
3. HIV infected individuals may develop AIDS, but may also experience a switch to SI phenotype. If the SI switch occurs first, it may change the risk to progress to AIDS.

The last of these is an example of a special class of multi-state models, called *illness-death models*. In this class of models individuals start out as healthy; this initial state will be denoted by state 1. They may become ill (move to state 2) and afterwards they may die (state 3). In principle they may also recover from their illness and become healthy again, i.e. move back to state 1. If this is possible the model is called a *bi-directional* illness-death model. Individuals may also die without first becoming ill (this is a direct transition from state 1 to state 3). A uni-directional illness-death model is illustrated in Figure 11.

Although, as suggested by the name, the typical application of an illness-death model is one where 'illness' is an unfavourable intermediate event, this is not necessarily the case. In Sections 4.4 and 4.5 we will use an illness-death model in bone marrow transplantation for illustration, where the 'illness' state corresponds to platelet recovery and 'death' corresponds to relapse or death.

Data on state occupation are often incomplete to some extent. States may be determined by the value of some marker that is not observed directly. Then the state can only be determined at the

DOI: 10.1002/sim

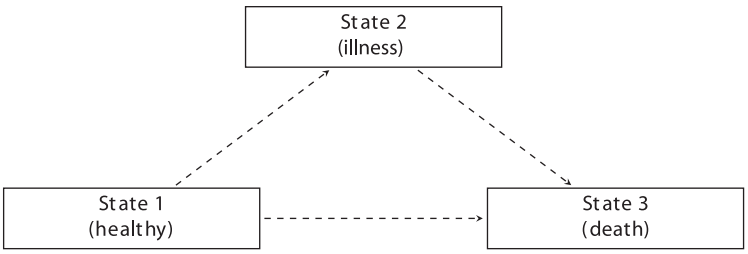

Figure 11. The illness-death model.

times at which the marker is measured and the transition time is interval censored. For a survey on multi-state models with interval censored data, see Reference [40]. Also, the marker value may be measured with error, leading to state misclassification. Models for these type of data will not be considered here, although software to fit these models will be mentioned in Section 4.7.

We will restrict attention in this tutorial to inference in multi-state models with non-parametric hazards in the framework of the Cox model, ignoring fully parametric and other important approaches, such as those based on additive hazards [41–43]. In Section 4.2 we introduce notation and discuss some preliminary notions. Section 4.3 is devoted to ways of representing data for multi-state modelling. Section 4.4 concerns estimation of regression coefficients and survival functions in multi-state models. Finally, Section 4.5 shows how to use multi-state models for prediction.

4.2. Preliminaries

4.2.1. Notation. We restrict to uni-directional multi-state models without recurrent events for which the intermediate transition times are observed exactly. Typically, a multi-state model contains one initial state, which we will assign the number 1. In the above examples, this state is entered at the moment of surgery for cancer, bone marrow transplantation and HIV infection respectively. Some states represent an endpoint; when a patient enters such a state, he or she will remain there or one is not interested in what happens after this state has been reached. We call these states *final* or *absorbing* states (the latter terminology comes from the theory of Markov chains and processes [44]). The absorbing states in our examples are death (in the cancer example), relapse and death (BMT), AIDS (HIV/AIDS). States that are neither initial nor absorbing states are called *intermediate* or *transient* states (again borrowed from Markov chain theory); strictly speaking, the initial state is also transient.

In Figures 10 or 11, each state is represented by a box. Transitions are represented by arrows going from one state to another. When we assign numbers to all states, we represent a transition from state i to j by '$i \to j$'. If T denotes the time of reaching state j from state i, we denote the hazard rate (transition intensity) of the $i \to j$ transition by (cf. (1) and (10))

$$\lambda_{ij}(t) = \lim_{\Delta t \downarrow 0} \frac{\text{Prob}(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t} \tag{22}$$

Similar to (3), we define the cumulative hazard for transition $i \to j$ by

$$\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)\, \mathrm{d}s \tag{23}$$

DOI: 10.1002/sim

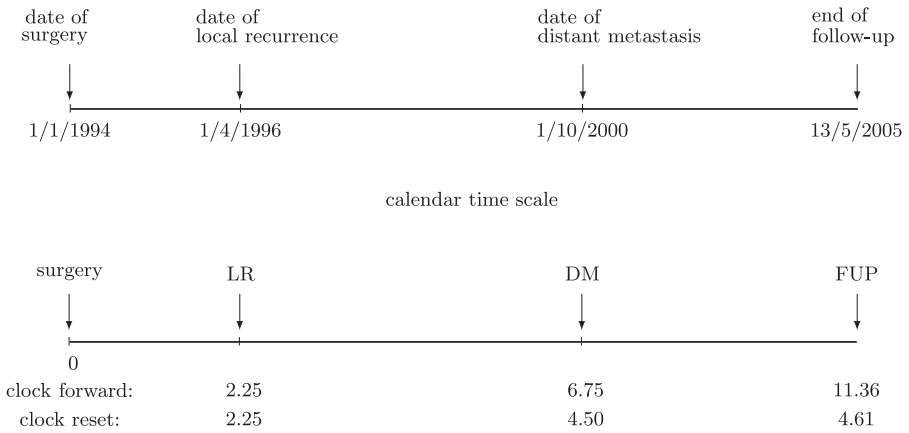

Figure 12. Illustration of the 'clock forward' and 'clock reset' approach. LR, DM and FUP stand for local recurrence, distant metastasis and follow-up, respectively.

4.2.2. Time scales. In the above definition, the question remains: what is t, or more precisely, what is the time scale to which t refers? Two approaches are in frequent use, which we shall denote here by the 'clock forward' or 'clock reset' approach.

'*Clock forward*': Time t refers to the time since the patient entered the initial state. The clock keeps moving forward for the patient, also when intermediate events occur.

'*Clock reset*': Time t in $\lambda_{ij}(t)$ refers to the time since entry in state i, also called backward recurrence time. The clock is reset to 0 each time the patient enters a new state.

The difference between the two approaches is illustrated in Figure 12. The upper half shows the dates of surgery and subsequent events for a cancer patient. At 13 May 2005, the patient is still alive. The lower picture shows the patient time-scale, first in the 'clock forward' approach, where time is measured from date of surgery, then in the 'clock reset' approach, where time intervals between state visits are recorded. In both instances the patient is censored for the last event, due to the end of follow-up.

4.2.3. Markov, semi-Markov and extended Markov models. A property that is often assumed in practice is that the multi-state model is a Markov model. Loosely speaking, the Markov property states that the future depends on the history only through the present. For a multi-state model this means that, given the present state and the event history of a patient, the next state to be visited and the time at which this will occur will only depend on the present state. Strictly speaking, only 'clock forward' models can be Markov models; for 'clock reset' models the Markov property cannot hold since the time scale itself depends on the history through the time since the current state was reached. However, if it is assumed that the sojourn times depend on the history of the process only through the present state and the time since entry of that state, the resulting multi-state model forms a sequence of embedded Markov models, called a Markov renewal model (see e.g. References [45–48]), or also a *semi-Markov* model. Note that competing risks models are always Markovian, since there is no event history.

Several kinds of violations (or relaxations) of the Markov property can be envisaged. One is a situation where the order of states visited influences transition rates. For example, in the multi-state

DOI: 10.1002/sim

model of Figure 10, the transition rate from local recurrence and distant metastasis to death can be different according to whether local recurrence was diagnosed before or after distant metastasis. Often in such situations, the multi-state model can be adapted (in this case by allowing two states, namely 'LR, then DM', and 'DM, then LR' to represent the 'local recurrence and distant metastasis' state) so that the multi-state model becomes Markov again. A second, more common relaxation of the Markov assumption is to let the sojourn times as covariates depend on the times at which earlier states have been entered. In the illustration in Sections 4.4 and 4.5 of this paper, we shall use the term *state arrival extended* (*semi-*)*Markov* to mean that the $i \to j$ transition hazard depends on the time of arrival at state i. For estimation in our illness-death model, the (semi-)Markov model is extended with only one additional parameter, associated with the arrival time at state 2 (or possibly a function of it), for the $2 \to 3$ transition only.

4.3. Data preparation

Many survival studies have their data stored initially in a one-row-per-subject (wide) format. This way is most convenient for most standard survival analyzes involving one endpoint. For example, consider the following three patients from a breast cancer study:

```
  patid survyrs survstat lryrs lrstat dmyrs dmstat
1     1    8.70        0  8.70      0  8.70      0
2     2    6.30        1  6.30      0  6.30      0
3     3   11.36        0  2.25      1  6.75      1
```

Patient 1 did not experience any event, i.e. is alive and event-free at $t = 8.7$ years. Patient 2 died after 6.3 years without any intermediate event. Patient 3 is as illustrated in Figure 12; she experienced a local recurrence at 2.25 years post-surgery, subsequently a distant metastasis at 6.75 years post-surgery, and is still alive at 11.36 years post-surgery.

The format allowing most flexibility for multi-state modelling is the so-called long format, already mentioned in Section 3. Each row now represents one patient 'at risk' for a certain transition. For the above 3 patients, for the multi-state model of Figure 10, the data in this format would need 12 rows instead of 3:

```
   patid  start   stop  status  from  to  transition  time
1      1   0.00   8.70       0     1   2       1 -> 2  8.70
2      1   0.00   8.70       0     1   3       1 -> 3  8.70
3      1   0.00   8.70       0     1   5       1 -> 5  8.70
4      2   0.00   6.30       0     1   2       1 -> 2  6.30
5      2   0.00   6.30       0     1   3       1 -> 3  6.30
6      2   0.00   6.30       1     1   5       1 -> 5  6.30
7      3   0.00   2.25       1     1   2       1 -> 2  2.25
8      3   0.00   2.25       0     1   3       1 -> 3  2.25
9      3   0.00   2.25       0     1   5       1 -> 5  2.25
10     3   2.25   6.75       1     2   4       2 -> 4  4.50
11     3   2.25   6.75       0     2   5       2 -> 5  4.50
12     3   6.75  11.36       0     4   5       4 -> 5  4.61
```

The data contains a patient identification column `patid` and a `transition` column, as well as a `from` and a `to` column specifying *from* which state the transition initiates and *to* which it

DOI: 10.1002/sim

terminates. Furthermore, it contains a `start` and `stop` time to indicate when the patient started and stopped being at risk for that transition, and a `status` to denote whether or not (1 and 0, respectively) the patient reached the `to` state. Patients 1 and 2 are represented by three columns each, one for each of the transitions going out from state 0. Patient 1 has `status` $= 0$ for each of these transitions, patient 2 has `status` $= 1$ only for the $1 \to 5$ transition (surgery to death). Patient 3 has these same three initial rows as well. After a local recurrence (`status` $= 1$ for the $1 \to 2$ transition), two more rows are added, corresponding to the two transitions ($2 \to 4$ and $2 \to 5$) going out from state 2. The `start` time for these transitions is 2.25, the `stop` time is 6.75. This is an example of delayed entry or left truncation (Section 2); patient 3 becomes at risk for the transitions $2 \to 4$ and $2 \to 5$ after 2.25 years. The variable `status` has value 0 for the $2 \to 5$ transition and 1 for the $2 \to 4$ transition. One final row is added after the patient has reached state 4 (local recurrence and distant metastasis). The only transition going out from state 4 is the $4 \to 5$ transition. The `start` time is 6.75, `stop` time is 11.36 years, the end of follow-up of that patient. Since the patient is still alive (censored), `status` $= 0$ for that row. One column `time` is added for modelling the 'clock reset' approach; it is simply defined by `time` = `stop`–`start`. If time-fixed covariates are also recorded, the values are simply replicated for each row corresponding to the same patient.

4.4. Estimation

We will illustrate estimation of the effect of prognostic factors on the transition rates in multi-state models, using the simplest non-trivial multi-state model, the illness-death model. Some aspects that play a role and that we will try to cover here are:

- which baseline hazards (for the different transitions) to choose proportional;
- whether to use the 'clock forward' or 'clock reset' approach;
- whether to use a (semi-)Markov or a state arrival extended (semi-)Markov model.

We will use data from the European Blood and Marrow Transplant registry (EBMT) for illustration in this and the next subsection. The data consists of 2204 patients in this registry, who received bone marrow transplantation between 1995 and 1998, and who had complete information on the prognostic factors considered here. These are as summarized in Table II.

Table II. Prognostic factors for all patients.

Prognostic factor		*n*	(%)
Disease classification	AML*	853	(39)
	ALL	447	(20)
	CML	904	(41)
Donor recipient	No gender mismatch*	1648	(75)
	Gender mismatch	556	(25)
GvHD prevention	No T-cell depletion (−TCD)*	1928	(87)
	+ TCD	276	(13)
Age at transplant (years)	$\leqslant 20$*	419	(19)
	20–40	1057	(48)
	>40	728	(33)

*Refers to reference category.

DOI: 10.1002/sim

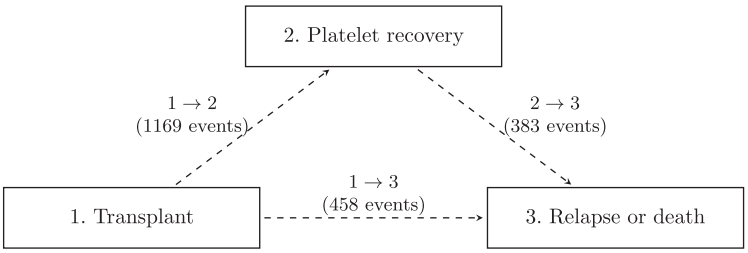

Figure 13. The EBMT illness-death model.

The multi-state model that we shall use for illustration here and in the next subsection is the bone marrow transplantation illness-death model already mentioned in Section 4.1. Here, the 'illness' state corresponds to platelet recovery and 'death' corresponds to relapse or death. The model is illustrated in Figure 13 along with the number of events. We can see that for 1169 of 2204 patients (53 per cent), platelet levels returned to normal levels; 383 of these 1169 (33 per cent) subsequently relapsed or died, the remaining 786 (67 per cent) did not relapse or die after platelet recovery. There were 458 patients (21 per cent) that relapsed or died without platelet recovery prior to relapse or death. Finally, 577 (26 per cent) of all 2204 patients did not experience any event in our data.

Let us start by not assuming anything about the baseline hazards. We will take the 'clock forward' approach and assume a Markov model. We will use Cox's proportional hazards model for each of the transition hazards separately. The hazard for transition $i \to j$ for a subject with covariate vector $\mathbf{Z}$ is then given by

$$\lambda_{ij}(t|\mathbf{Z}) = \lambda_{ij,0}(t)\exp(\boldsymbol{\beta}_{ij}^{\top}\mathbf{Z}) \tag{24}$$

where $\lambda_{ij,0}(t)$ is the baseline hazard of transition $i \to j$, and $\boldsymbol{\beta}_{ij}$ is the vector of regression coefficients that describe the effect of $\mathbf{Z}$ on transition $i \to j$. For estimation in the $1 \to 2$ ($1 \to 3$) transition, in long format, it suffices to select only the rows corresponding to `transition = 1->2` (`transition = 1->3`), and use a Cox regression on the selected data. For estimation of regression parameters for the $2 \to 3$ transition (platelet recovery $\to$ relapse or death), it is important to realise that patients are at risk only after entering state 2 (delayed entry). The estimates of $\boldsymbol{\beta}_{ij}$, their standard errors and P-values are reported in Table III in the *Markov stratified hazards* column. The most important findings are the higher relapse/death rates for older patients (particularly older than 40), both before and after platelet recovery, the lower platelet recovery rate for CML patients, and the increased platelet recovery rate for patients receiving T-cell depletion.

The estimated cumulative baseline hazards (i.e. all covariate values equal to the reference value) for this model are shown in the left plot of Figure 14. Note the sharp increase in the baseline rate of platelet recovery.

We may assume the baseline hazards of the $1 \to 3$ transition and the $2 \to 3$ transitions to be proportional. In view of Figure 14 this does not seem unreasonable. This is equivalent to grouping the $1 \to 3$ and $2 \to 3$ transitions and using the occurrence of the intermediate event as a time-dependent covariate. The model for the transition to state 3 is then given by

$$\begin{aligned}\lambda_{13}(t|\mathbf{Z}) &= \lambda_{3,0}(t)\exp(\boldsymbol{\beta}_{13}^{\top}\mathbf{Z})\\ \lambda_{23}(t|\mathbf{Z}) &= \lambda_{3,0}(t)\exp(\boldsymbol{\beta}_{23}^{\top}\mathbf{Z} + \delta)\end{aligned} \tag{25}$$

DOI: 10.1002/sim

Table III. Parameter estimates in different models; 'clock forward' approach.

		Markov				State arrival extended Markov proportional hazards	
		Stratified hazards		Proportional hazards			
		coef (SE)	*P*	coef (SE)	*P*	coef (SE)	*P*
$1 \rightarrow 2$ transition							
Disease classification	AML						
	ALL	−0.044 (0.078)	0.58				
	CML	−0.297 (0.068)	<0.0001				
Age at	⩽20						
transplantation	20–40	−0.165 (0.079)	0.037				
	>40	−0.090 (0.086)	0.30				
Donor recipient	No gender mism.						
	Gender mismatch	0.046 (0.067)	0.49				
GvHD prevention	No TCD						
	+ TCD	0.429 (0.080)	<0.0001				
$1 \rightarrow 3$ transition							
Disease classification	AML						
	ALL	0.256 (0.135)	0.058	0.261 (0.135)	0.054	0.261 (0.135)	0.054
	CML	0.017 (0.108)	0.88	0.004 (0.108)	0.97	0.004 (0.108)	0.97
Age at	⩽20						
transplantation	20–40	0.255 (0.151)	0.091	0.251 (0.151)	0.097	0.251 (0.151)	0.097
	>40	0.526 (0.158)	0.0009	0.526 (0.158)	0.0009	0.526 (0.158)	0.0009
Donor recipient	No gender mism.						
	Gender mismatch	−0.075 (0.110)	0.50	−0.072 (0.110)	0.51	−0.072 (0.110)	0.51
GvHD prevention	No TCD						
	+ TCD	0.297 (0.150)	0.048	0.319 (0.150)	0.034	0.319 (0.150)	0.034
Disease classification	AML						
	ALL	0.136 (0.148)	0.36	0.140 (0.148)	0.34	0.132 (0.149)	0.38
	CML	0.247 (0.117)	0.035	0.250 (0.117)	0.032	0.252 (0.117)	0.031
Age at	⩽20						
transplantation	20–40	0.062 (0.153)	0.69	0.056 (0.153)	0.72	0.058 (0.153)	0.70
	>40	0.581 (0.160)	0.0003	0.562 (0.160)	0.0004	0.566 (0.160)	0.0004
Donor recipient	No gender mism.						
	Gender mismatch	0.173 (0.115)	0.13	0.169 (0.114)	0.14	0.167 (0.115)	0.15
GvHD prevention	No TCD						
	+ TCD	0.201 (0.126)	0.11	0.211 (0.126)	0.094	0.207 (0.126)	0.10
Platelet recovery		NA	NA	−0.379 (0.212)	0.073	−0.407 (0.219)	0.063
Time of platelet recovery		NA	NA	NA	NA	0.295 (0.595)	0.62

DOI: 10.1002/sim

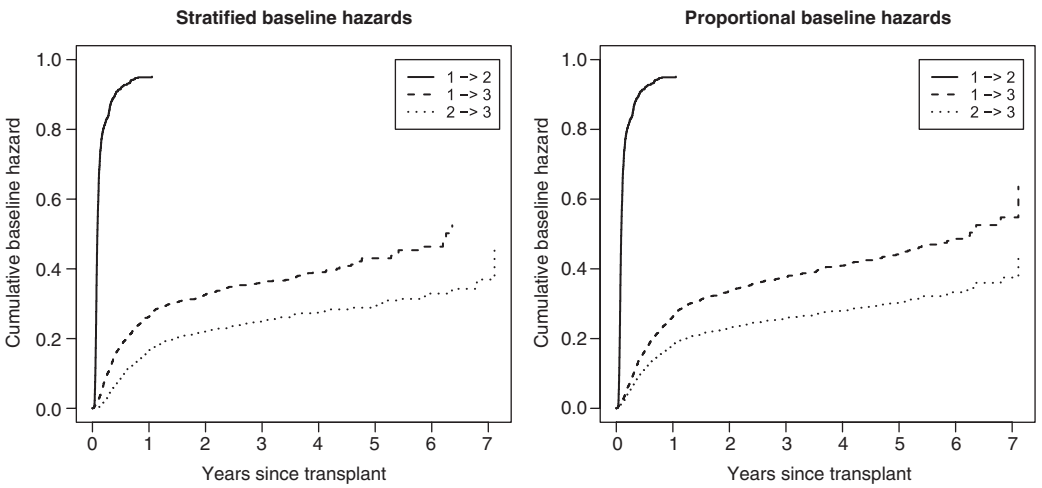

Figure 14. Baseline cumulative hazard curves for the EBMT illness-death model. On the left the baseline hazards are not assumed to be proportional; on the right the baseline hazards of the $1 \to 3$ and the $2 \to 3$ transitions are related through a proportional hazards assumption on the transition hazards (see Equation (25)).

Note the use of the same baseline hazard $\lambda_{3,0}(t)$ for both the $1 \to 3$ and the $2 \to 3$ transition. The regression coefficients $\boldsymbol{\beta}_{13}$ and $\boldsymbol{\beta}_{23}$ in (25) have the same interpretation as in (24) (but will typically result in different though comparable estimates). Another way of expressing (25) is as

$$\lambda_{j3}(t) = \lambda_{3,0}(t) \exp(\boldsymbol{\beta}_{13}^{\top}\mathbf{Z} + (j-1)\boldsymbol{\Delta}^{\top}\mathbf{Z} + (j-1)\delta), \quad j = 1, 2 \tag{26}$$

i.e. as a single Cox regression with main covariate effects and interactions with transition. Here $\boldsymbol{\beta}_{13}$ still represents the covariate effects for relapse or death before platelet recovery; $\boldsymbol{\Delta}$ represents the difference in covariate effects for relapse or death after platelet recovery, compared to before platelet recovery. Thus, $\boldsymbol{\beta}_{23}$ can be retrieved from $\boldsymbol{\beta}_{13}$ and $\boldsymbol{\Delta}$ by the simple relationship $\boldsymbol{\beta}_{23} = \boldsymbol{\beta}_{13} + \boldsymbol{\Delta}$. The results are also found in Table III, in column *Markov proportional hazards*. The estimates of $\boldsymbol{\beta}_{12}$ are left blank; they are the same as before. None of the elements of $\boldsymbol{\Delta}$ was found to be significantly different from zero (not shown). There are a number of advantages to taking the baseline hazards of $1 \to 3$ and $2 \to 3$ proportional. The fact that the same baseline hazard $\lambda_{3,0}(t)$ is used for both the $1 \to 3$ and the $2 \to 3$ transition may in case of rare events result in higher precision of the estimates of $\boldsymbol{\beta}_{13}$ and/or $\boldsymbol{\beta}_{23}$. The model can be fitted as a time-dependent Cox model, where the intermediate state is added as a time-dependent covariate. Model (26) could be stated in this form as

$$\lambda_3(t) = \lambda_{3,0}(t) \exp(\boldsymbol{\beta}_{13}^{\top}\mathbf{Z} + \mathrm{PR}(t)\boldsymbol{\Delta}^{\top}\mathbf{Z} + \mathrm{PR}(t)\delta)$$

where the time-dependent covariate $\mathrm{PR}(t)$ equals 0 for all t before time of platelet recovery, and 1 after time of platelet recovery. More importantly, the parameter δ gives useful additional information. Its exponent $\exp(\delta)$ represents the effect of experiencing the intermediate event on the rate of occurrence of the endpoint. Here the parameter δ is estimated as -0.38 with a standard error of 0.21 (Table III; $P = 0.073$). This means that platelet recovery has a (trend significant) protective effect on relapse-free survival (hazard ratio $= 0.67$, 95 per cent CI: 0.43–1.02). On the

DOI: 10.1002/sim

downside of course, one then has to wonder whether the assumption of proportional hazards for the two transitions is reasonably fulfilled. The right plot of Figure 14 shows the estimated baseline cumulative hazards for the three transitions in this model. Comparison with the left plot of Figure 14 strongly suggests that the proportionality assumption of the baseline hazards for the $1 \to 3$ and $2 \to 3$ transitions is reasonably fulfilled. This can be checked more rigorously by testing the significance of an interaction between (a function of) time, $f(t)$, and an indicator distinguishing between the two transitions, or by checking the Schoenfeld residuals. We will not pursue this here.

The last two columns of Table III contain parameter estimates and P-values for the state arrival extended Markov model. We have assumed proportionality of the baseline hazards for the $1 \to 3$ and $2 \to 3$ transitions. The model was fitted by adding the extra term pr in (26), where pr is the time (in years) from transplant to platelet recovery. This term is meaningful only for the $2 \to 3$ transition (and is therefore set to 0 for the $1 \to 3$ transition).

Table IV contains the results for the 'clock reset' approach, for the same three models. The results for the $1 \to 2$ transition are omitted here, since they are again identical to those in Table III. The difference in parameter estimates between the 'clock forward' and the 'clock reset' approaches are small; most notable are the differences in the effects of platelet recovery and the time at which it occurs on the relapse-free survival after platelet recovery.

It is hard to give general guidance in deciding between the 'clock forward' and 'clock reset' approaches. The clinical context will most often be the most important consideration here. In our experience the difference between the two approaches is usually quite small with regard to the estimated regression coefficients. Farewell and Cox [49] propose a more formal procedure in studying which time-scale is most appropriate. For the remainder of this tutorial, we will use the (clock-forward) Markov proportional hazards model.

4.5. Prediction

In the preceding subsection, we have modelled the effects of covariates on the transition hazard. In Section 3 on competing risks we have already seen that effects on the cumulative incidence function may be different from what the regression coefficients suggest. In a multi-state setting, this becomes even more of an issue, since intermediate events also contribute to effects on the cumulative scale. This subsection is devoted to estimation of cumulative effects, or prediction, to answer clinically important questions such as the following in our example:

- Given a bone marrow transplantation patient whose platelets have recovered after 60 days and who has had no further events at one year post-transplant, what is then the probability of surviving relapse-free for 2 more years? How does this probability compare to a patient whose platelets have not yet recovered?

The general problem is to estimate the conditional probabilities of some clinical future events, given an (event) history, and possibly a set of values for prognostic factors $\mathbf{Z}$ of a patient. The estimates of these probabilities are based on the results obtained from the Cox model on the transition hazards between the states. Let u be the time at which the prediction is made measured from the time origin of the patient (surgery, transplantation, HIV infection in the context of the three examples mentioned in Section 4.1). Let us also denote the event history of the patient by $\mathscr{H}_u$. This event history contains the times of all events recorded for that patient and the event types. Let E_t denote some future event evaluated at time t, e.g. the event of surviving relapse-free until $t = 10$ years

DOI: 10.1002/sim

Table IV. Parameter estimates in different models; 'clock reset' approach.

		Semi-Markov				State arrival extended Semi-Markov proportional hazards	
		Stratified hazards		Proportional hazards			
		coef (SE)	P	coef (SE)	P	coef (SE)	P
$1 \rightarrow 3$ transition							
Disease	AML						
classification	ALL	0.256 (0.135)	0.058	0.259 (0.135)	0.056	0.259 (0.135)	0.056
	CML	0.017 (0.108)	0.88	0.008 (0.108)	0.94	0.008 (0.108)	0.94
Age at	$\leqslant 20$						
transplantation	20–40	0.255 (0.151)	0.091	0.252 (0.151)	0.095	0.252 (0.151)	0.095
	>40	0.526 (0.158)	0.0009	0.528 (0.158)	0.0008	0.528 (0.158)	0.0008
Donor recipient	No gender mism.						
	Gender mismatch	−0.075 (0.110)	0.50	−0.073 (0.110)	0.51	−0.073 (0.110)	0.51
GvHD prevention	No TCD						
	+ TCD	0.297 (0.150)	0.048	0.310 (0.150)	0.039	0.310 (0.150)	0.039
$2 \rightarrow 3$ transition							
Disease	AML						
classification	ALL	0.120 (0.148)	0.42	0.117 (0.148)	0.43	0.138 (0.149)	0.36
	CML	0.252 (0.117)	0.031	0.253 (0.117)	0.030	0.249 (0.117)	0.033
Age at	$\leqslant 20$						
transplantation	20–40	0.065 (0.153)	0.67	0.064 (0.153)	0.68	0.058 (0.153)	0.70
	>40	0.582 (0.160)	0.0003	0.574 (0.160)	0.0003	0.567 (0.160)	0.0004
Donor recipient	No gender mism.						
	Gender mismatch	0.170 (0.115)	0.14	0.164 (0.114)	0.15	0.170 (0.115)	0.14
GvHD prevention	No TCD						
	+ TCD	0.197 (0.126)	0.12	0.201 (0.126)	0.11	0.209 (0.126)	0.098
Platelet recovery		NA	NA	−0.416 (0.211)	0.049	−0.350 (0.219)	0.11
Time of platelet recovery		NA	NA	NA	NA	−0.658 (0.595)	0.26

DOI: 10.1002/sim

after study entry. Then we are interested in

$$\text{Prob}(E_t|\mathscr{H}_u, \mathbf{Z})$$

Given a multi-state model without recurrent events, these probabilities can be expressed in terms of the hazards for the transitions and can be estimated by appropriately combining the estimated baseline hazards and regression coefficients. This was first outlined in Reference [36] using work of Arjas and Eerola [50].

We will derive the formulas and show how they can be estimated for the illness-death model, more specifically the bone marrow transplantation model of Figure 13. We will first illustrate the 'clock forward' approach. We denote the time of the intermediate event by R and the time of the final event by T. Let us start with prediction from state 2, i.e. a patient whose platelets have recovered at time $R=r$ after transplant. We denote the corresponding event history by

$$\mathscr{H}_{2,r}(u)=\{R=r, T>u\} \quad \text{with } r\leqslant u$$

In the notation, u is the time of prediction (measured from the time origin), the subscript 2 denotes the fact that the patient is in state 2 at time u, and r is the time of reaching state 2, i.e. the time of the intermediate event or illness, in this instance platelet recovery. The probabilities

$$P_{23,r}(u,t)=\text{Prob}(T\leqslant t|\mathscr{H}_{2,r}(u)), \quad P_{22,r}(u,t)=\text{Prob}(T>t|\mathscr{H}_{2,r}(u)) \tag{27}$$

are the (conditional) probabilities of going to state 3 before or at time t and staying in state 2 until time t, respectively, given a recovery of platelets at time r and no further events until time of prediction u. The notation is similar to that of the histories; in addition t is time (since patient entry) for which the prediction probability has to be calculated. The probabilities in (27) can be directly expressed in terms of the hazard rate for transition $2\to 3$, for instance

$$P_{23,r}(u,t)=\int_u^t \lambda_{23,r}(s)\exp\left(-\int_u^s \lambda_{23,r}(v)\,\mathrm{d}v\right)\mathrm{d}s \tag{28}$$

An explanation of this formula is as follows. In order for the patient, starting from state 2 at time u, to be in state 3 at time t, he or she as to make the transition from 2 to 3 at some time s (factor $\lambda_{23,r}(s)$). Up to time s, the patient has to remain in state 2 (factor $\exp(-\int_u^s \lambda_{23,r}(v)\,\mathrm{d}v)$). Using $S_{2,r}(s)=\exp(-\int_0^s \lambda_{23,r}(v)\,\mathrm{d}v)$, we can simplify (28) to

$$P_{23,r}(u,t)=\int_u^t \lambda_{23,r}(s)S_{2,r}(s)\,\mathrm{d}s/S_{2,r}(u)=\frac{S_{2,r}(u)-S_{2,r}(t)}{S_{2,r}(u)}=1-\frac{S_{2,r}(t)}{S_{2,r}(u)} \tag{29}$$

Since there is only one state to reach from state 2, we have

$$P_{22,r}(u,t)=1-P_{23,r}(u,t)=\frac{S_{2,r}(t)}{S_{2,r}(u)}$$

In a Markov model, $P_{22,r}(u,t)$ and $P_{23,r}(u,t)$ do not depend on r because $\lambda_{23,r}(t)$ does not depend on the time r of 'illness'.

For a patient who is alive and relapse-free and whose platelets have not (yet) recovered at time u (i.e. who is in state 1 at time u), we denote the corresponding history by

$$\mathscr{H}_1(u)=\{R>u, T>u\}$$

DOI: 10.1002/sim

We can now discern four different scenarios for a patient with history $\mathscr{H}_1$; their probabilities are denoted by

$$\begin{aligned} P_{11}(u,t) &= \text{Prob}(R>t, T>t|\mathscr{H}_1(u)) \\ P_{12}(u,t) &= \text{Prob}(R\leqslant t, T>t|\mathscr{H}_1(u)) \\ P^1_{13}(u,t) &= \text{Prob}(T\leqslant t, T<R|\mathscr{H}_1(u)) \\ P^2_{13}(u,t) &= \text{Prob}(R\leqslant T\leqslant t|\mathscr{H}_1(u)) \end{aligned} \tag{30}$$

The superscripts in P^1_{13} and P^2_{13} serve to distinguish between relapse or death without platelet recovery (directly from state 1) and after platelet recovery (moving through state 2), respectively. The probability of relapse or death before time t, conditionally given no events at time u, $P_{13}(u,t) = \text{Prob}(T\leqslant t|\mathscr{H}_1(u))$ is the sum of the corresponding probabilities without and with prior platelet recovery, $P^1_{13}(u,t) + P^2_{13}(u,t)$.

Let us start with the most complicated of these probabilities. For $t>u$, we have

$$\begin{aligned} P^2_{13}(u,t) &= \int_u^t \lambda_{12}(r) \exp\left(-\int_u^r (\lambda_{12}(v)+\lambda_{13}(v))\,\mathrm{d}v\right) P_{23,r}(r,t)\,\mathrm{d}r \\ &= \frac{\int_u^t \lambda_{12}(r) S_1(r) P_{23,r}(r,t)\,\mathrm{d}r}{S_1(u)} \end{aligned} \tag{31}$$

The explanation of this formula is similar to that of (29). Given that the patient starts in state 1 at time u, in order to first visit state 2 and then state 3 before or at time t, the patient must at some time r between u and t visit state 2, before time r not having visited states 2 or 3. This is expressed by $\lambda_{12}(r)\exp(-\int_u^r(\lambda_{12}(v)+\lambda_{13}(v))\,\mathrm{d}v)$ (cf. equations (11) and (12)). Once the patient has reached state 2 at time r, the probability of reaching state 3 before or at time t is given by $P_{23,r}(r,t)$ as above. The second equation in (31) follows by defining

$$S_1(t) = \exp(-(\Lambda_{12}(t)+\Lambda_{13}(t))), \quad \Lambda_{1j}(t) = \int_0^t \lambda_{1j}(s)\,\mathrm{d}s$$

as the probability of staying in state 1 until time t. Note the similarity with the competing risks context (in particular, equation (11)). Similarly, again for $t>u$, we also have

$$P_{12}(u,t) = \frac{\int_u^t \lambda_{12}(r) S_1(r) P_{22,r}(r,t)\,\mathrm{d}r}{S_1(u)} \tag{32}$$

i.e. equation (31) with $P_{23,r}(r,t)$ replaced by $P_{22,r}(r,t)$. The sum of (31) and (32),

$$P^2_{13}(u,t) + P_{12}(u,t) = \frac{\int_u^t \lambda_{12}(r) S_1(r)\,\mathrm{d}r}{S_1(u)} \tag{33}$$

is the conditional probability of having visited state 2 by time t, given $\mathscr{H}_1(u)$. Evaluated at $u=0$, this is the cumulative incidence function of state 2, cf. equation (12). The conditional probability

DOI: 10.1002/sim

of having visited state 3 by time t without moving through state 2, given $\mathscr{H}_1(u)$, $P_{13}^1(u,t)$, has a similar expression, with λ_{12} replaced by λ_{13}:

$$P_{13}^1(u,t) = \frac{\int_u^t \lambda_{13}(r) S_1(r)\,\mathrm{d}r}{S_1(u)} \tag{34}$$

The sum of the probabilities in (33) and (34),

$$\begin{aligned} P_{13}^1(u,t) + P_{12}(u,t) + P_{13}^2(u,t) &= \frac{\int_u^t (\lambda_{12}(r) + \lambda_{13}(r)) S_1(r)\,\mathrm{d}r}{S_1(u)} \\ &= \frac{S_1(u) - S_1(t)}{S_1(u)} = 1 - \frac{S_1(t)}{S_1(u)} \end{aligned}$$

is the conditional probability of having left state 1 before or at time t, given $\mathscr{H}_1(u)$. Its complement, the conditional probability of staying in state 1 until time t, given $\mathscr{H}_1(u)$, is

$$P_{11}(u,t) = \frac{S_1(t)}{S_1(u)} \tag{35}$$

Similar formulas can also be derived for the 'clock reset' approach. In general, they are slightly more complicated because of the presence of several time-scales. However, in the easier case of the illness-death model, the only difference is in the prediction probabilities out of state 2 given $\mathscr{H}_{2,r}(u)$. Instead of (29), these now become

$$P_{22,r}(u,t) = \frac{S_{2,r}(t-r)}{S_{2,r}(u-r)}, \quad P_{23,r}(u,t) = 1 - \frac{S_{2,r}(t-r)}{S_{2,r}(u-r)} \tag{36}$$

In the special case where $u = r$ (these are the probabilities used in (31) and (32)), (36) simplifies to

$$P_{22,r}(r,t) = S_{2,r}(t-r), \quad P_{23,r}(u,t) = 1 - S_{2,r}(t-r) \tag{37}$$

since for $u = r$, $S_{2,r}(u-r) = 1$. It is important to note that estimators of $S_{2,r}$ will also change compared to the 'clock forward' approach, due to the different time-scale. For formulas in more extensive multi-state models, see Reference [36] for the 'clock forward' and References [35, 48] for the 'clock reset' approach.

How do we estimate these prediction probabilities from data? Given a multi-state model with estimated regression coefficients $\widehat{\boldsymbol{\beta}}_{ij}$ and baseline cumulative hazards $\widehat{\Lambda}_{ij,0}(t)$, we can compute the cumulative hazards $\widehat{\Lambda}_{ij}(t|\mathbf{Z}) = \widehat{\Lambda}_{ij,0}(t)\exp(\widehat{\boldsymbol{\beta}}_{ij}^\top \mathbf{Z})$ corresponding to a patient with covariate values $\mathbf{Z}$. This estimator will typically be represented by jumps of size $\widehat{\lambda}_{ij}(s)$ at certain time points s at which events for the $i \to j$ transition occur (or for which events for other transitions occur as well, depending on proportional hazards assumptions between baseline hazards from different transitions). We denote the collection of these time points by $\mathscr{T}_{ij}$. One then simply obtains estimates of the prediction probabilities in (31)–(35) and (36) by replacing each integral by a sum and by

DOI: 10.1002/sim

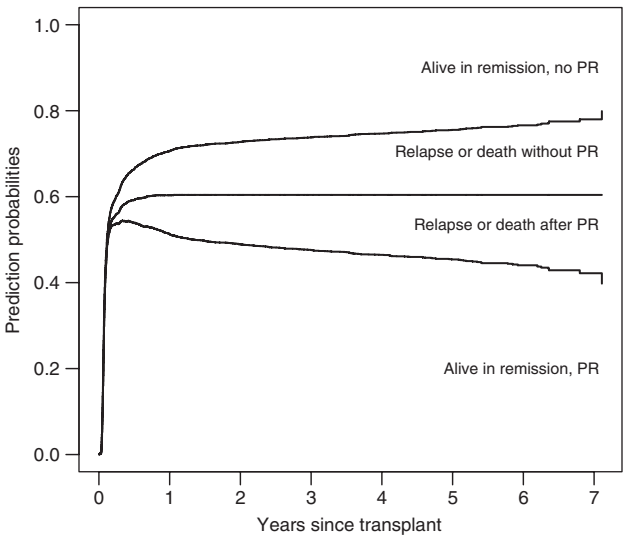

Figure 15. Stacked prediction probabilities at $u = 0$ for a reference patient. PR stands for platelet recovery.

replacing hazard and survival functions by their estimated counterparts. For instance,

$$\widehat{P}^2_{13}(u,t) = \frac{\sum_{\substack{r \in \mathscr{T}_{12} \\ 0 \leqslant r \leqslant t}} \widehat{\lambda}_{12}(r)\widehat{S}_1(r-)\widehat{P}_{23,r}(r,t)}{\widehat{S}_1(u)}$$

where

$$\widehat{S}_1(r) = \prod_{s \leqslant r} (1 - (\widehat{\lambda}_{12}(s) + \widehat{\lambda}_{13}(s)))$$

The latter equation is a discretized version of $S_1(r) = \exp(-(\Lambda_{12}(r) + \Lambda_{13}(r)))$, cf. (15). Note that these prediction probabilities are special cases of the Aalen–Johansen estimator [51] (for more details see Reference [2, Section IV.4]).

In the remainder of this subsection we shall illustrate the use of these prediction probabilities for the EBMT multi-state model of Figure 13, based on the 'clock forward' Markov proportional hazards model of the previous subsection. Figure 15 shows, from bottom to top, the probabilities $\widehat{P}_{12}(u,t)$, $\widehat{P}^2_{13}(u,t)$, $\widehat{P}^1_{13}(u,t)$ and $\widehat{P}_{11}(u,t)$, for a patient with reference values for all covariates, i.e. an AML patient, $\leqslant 20$ years, no gender mismatch, no T-cell depletion. The time of prediction here is $u = 0$, right after the transplant. The probabilities are stacked; the distance between two curves represents the probability, associated with the text in the figure.

Figure 16 is similar to Figure 15; the time of prediction is now $u = 0.5$ years after transplant. The estimated prediction probabilities $\widehat{P}_{12}(u,t)$ and $\widehat{P}^2_{13}(u,t)$ are now much smaller (in fact, $\widehat{P}^2_{13}(u,t)$ is hardly visible on the plot). This is because the $1 \to 2$ hazard is initially very high (Figure 14) and then decreases rapidly. Given that after 0.5 years no platelet recovery has occurred, platelet recovery is far less likely to occur later.

Summing $\widehat{P}^1_{13}(u,t)$ and $\widehat{P}^2_{13}(u,t)$ gives $\widehat{P}_{13}(u,t)$, the estimated conditional probability of death or relapse prior to time t, given alive without relapse at time u. Conditional relapse-free survival is

DOI: 10.1002/sim

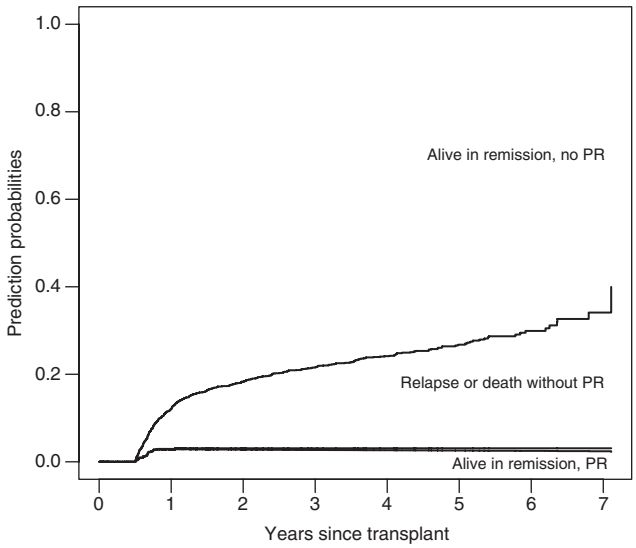

Figure 16. Stacked prediction probabilities at $u = 0.5$ years for a reference patient. The predicted probability of relapse or death after PR is negligible. PR stands for platelet recovery.

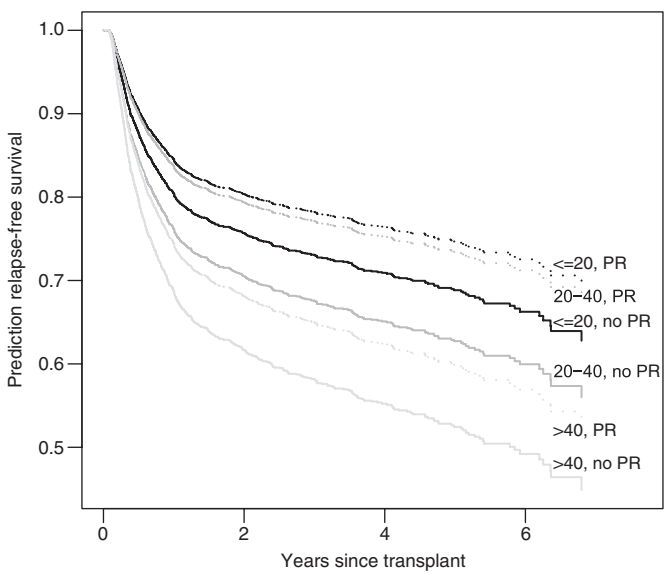

Figure 17. Predicted relapse-free survival probabilities for three patients in different age categories, given platelet recovery (dashed) and given no platelet recovery (solid). The time of prediction was 1 month after transplant. PR stands for platelet recovery.

estimated as $1 - \widehat{P}_{13}(u, t) = \widehat{P}_{11}(u, t) + \widehat{P}_{12}(u, t)$. Figure 17 shows predicted relapse-free survival probabilities for three patients, one in each age category. The other covariate values were set to their reference values. The time of prediction was one month after transplant. The solid lines are predicted

DOI: 10.1002/sim

relapse-free survival probabilities given no platelet recovery ($1-\widehat{P}_{13}(u,t)$), the dashed lines given platelet recovery ($1-\widehat{P}_{23}(u,t)$). Broadly summarizing, older patients have lower predicted relapse-free survival probabilities (worse prognosis), and the occurrence of platelet recovery improves prognosis. However, it can be seen that this improvement in prognosis is considerably higher, for instance, for the middle (20–40) age group than for the youngest ($\leqslant$20) age group. The reason for this is the fact that the effect of age 20–40 compared to age $\leqslant$20 is moderate (estimated coefficient = 0.251) for the direct transition of transplant to relapse or death ($1 \to 3$), while this effect is negligible (estimated coefficient = 0.056) for the transition from platelet recovery to relapse or death ($2 \to 3$).

4.6. Software

Just like in the competing risks situation, estimation of the transition intensities can be done in most statistical packages (for example S-plus/R, SAS, BMDP and Stata). Since SPSS does not allow for left truncation, it can only be used if we assume proportionality of the different transition hazards, such that a time-dependent Cox model is fitted. Estimation of cumulative effects is more complicated. Recently, an R package `changeLOS` [52] has become available that implements the Aalen–Johansen estimator for general multi-state models with non-parametric hazards. At the moment, it does not allow covariates. A package `mstate` for R is available from `http://www.msbi.nl/multistate` in which the data preparation and techniques used in this tutorial are implemented. This website also contains the EBMT data set used for illustration and the full code used to obtain all fitted models and prediction results.

4.7. Summary and concluding remarks

We have seen that estimation of transition intensities in Markov models and some of its extensions can be performed in a simple way. Estimation of cumulative effects is more complicated, due to the many possible pathways that may occur. Throughout, we assumed that all transitions were observed. However, irregular observation schemes may cause transition times to be interval censored, and transitions may even be missed. Moreover, especially if the states are determined by marker measurements, misclassification of the state may occur. Markov models that incorporate misclassification are called *hidden* Markov models. Markov models and hidden Markov models with interval censored transition times can be estimated via several stand alone programs and in R in the package `msm`. These models require the baseline hazard to be parametric (constant or piecewise constant).

ACKNOWLEDGEMENTS

This paper is based on a course given at the ISCB 2004 conference in Leiden. The authors thank the organizers for their invitation. Part of this work was supported by a grant (ZonMW 2002-912-02-015 Survival analysis for complicated data) from the Netherlands Organization for Scientific Research. The European Blood and Marrow Transplantation and the Amsterdam Cohort Studies on HIV Infection and AIDS are gratefully acknowledged for making data available for this tutorial.

REFERENCES

1. Andersen PK, Abildstrøm SZ, Rosthøj S. Competing risks as a multi-state model. *Statistical Methods in Medical Research* 2002; **11**:203–215.

DOI: 10.1002/sim

2. Andersen PK, Borgan Ø, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer: Berlin, 1993.
3. Hougaard P. *Analysis of Multivariate Survival Data*. Statistics for Biology and Health. Springer: New York, 2000.
4. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data* (2nd edn). Wiley: New York, 2002.
5. Klein JP, Moeschberger ML. *Survival Analysis. Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer: New York, 1997.
6. Marubini E, Valsecchi MG. *Analysing Survival Data from Clinical Trials and Observational Studies*. Wiley: New York, 2004.
7. Therneau TM, Grambsch PM. *Modeling Survival Data. Extending the Cox Model*. Statistics for Biology and Health. Springer: New York, 2000.
8. Andersen PK. Multi-state models. *Statistical Methods in Medical Research* 2002; **11**:89–90.
9. Andersen PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 2002; **11**:91–115.
10. R Development Core Team. *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2006. ISBN: 3-900051-07-0.
11. Hoover DR, Muñoz A, Carey V, Taylor JMG, Vanraden M, Chmiel JS, Kingsley L. Using events from dropouts in nonparametric survival function estimation with application to incubation of AIDS. *Journal of the American Statistical Association* 1993; **8**:37–43.
12. Bernoulli D. Essai d'une nouvelle analyse de la mortalité causée par la petite Vérole, et des avantages de l'inoculation pour la prevenir. *Mémoires de l'Académie Royal des Sciences*, *Paris* 1760; 1–45.
13. Gail M. A review and critique of some models used in competing risk analysis. *Biometrics* 1975; **31**:209–222.
14. Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics* 1978; **34**:541–554.
15. Geskus RB. On the inclusion of prevalent cases in HIV/AIDS natural history studies through a marker-based estimate of time since seroconversion. *Statistics in Medicine* 2000; **19**:1753–1769.
16. Geskus RB, Miedema FA, Goudsmit J, Reiss P, Schuitemaker H, Coutinho RA. Prediction of residual time to AIDS and death based on markers and cofactors. *Journal of AIDS* 2003; **32**:514–521.
17. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine* 1999; **18**:695–706.
18. Cox DR. The analysis of exponentially distributed lifetimes with 2 types of failure. *Journal of the Royal Statistical Society*, *Series B* 1959; **21**:411–421.
19. Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences U.S.A.*, 1975; **72**:20–22.
20. Gray RJ. A class of *k*-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* 1988; **16**:1141–1154.
21. van Rij RP, De Roda Husman AM, Brouwer M, Goudsmit J, Coutinho RA, Schuitemaker H. Role of CCR2 genotype in the clinical course of syncytium-inducing (SI) or non-SI human immunodeficiency virus type 1 infection and in the time to conversion to SI virus variants. *Journal of Infectious Diseases* 1998; **178**:1806–1811.
22. Andersen PK, Hansen LS, Keiding N. Nonparametric and semiparametric estimation of transition-probabilities from censored observation of a nonhomogeneous Markov process. *Scandinavian Journal of Statistics* 1991; **18**:153–167.
23. Holt JD. Competing risk analyses with special reference to matched pair experiments. *Biometrika* 1978; **65**: 159–165.
24. S original by Terry Therneau and ported by Thomas Lumley. *Survival*: *Survival Analysis*, *Including Penalised Likelihood*. R package Version 2.15.
25. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995; **51**:524–532.
26. Fiocco M, Putter H, van Houwelingen JC. Reduced rank proportional hazards model for competing risks. *Biostatistics* 2005; **6**:465–478.
27. Fiocco M, Putter H, van de Velde CJH, van Houwelingen JC. Reduced rank proportional hazards model for competing risks: an application. *Journal of Statistical Planning and Inference* 2006; **136**:1655–1668.
28. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**:496–509.
29. Rosthøj S, Andersen PK, Abildstrøm SZ. SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. *Computer Methods and Programs in Biomedicine* 2004; **74**:69–75.

DOI: 10.1002/sim

30. Di Serio C. The protective impact of a covariate on competing failures with an example from a bone marrow transplantation study. *Lifetime Data Analysis* 1997; **3**:99–122.
31. Andersen PK, Klein JP, Rosthøj S. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* 2003; **90**:15–27.
32. Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 2005; **61**:223–229.
33. Tai BC, White IR, Gebski V, Machin D. On the issue of 'multiple' first failures in competing risks analysis. *Statistics in Medicine* 2002; **21**:2243–2255.
34. Wohlfahrt J, Andersen PK, Melbye M. Multivariate competing risks. *Statistics in Medicine* 1999; **18**:1023–1030.
35. Putter H, van der Hage JA, de Bock GH, Elgalta R, van der Velde CJH. Estimation and prediction in a multistate model for breast cancer. *Biometrical Journal* 2006; **48**:366–380.
36. 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* 1994; **12**: 2315–2332.
37. Klein JP, Szydlo RM, Craddock C, Goldman JM. Estimation of current leukaemia-free survival following donor lymphocyte infusion therapy for patients with leukaemia who relapse after allografting: application of a multistate model. *Statistics in Medicine* 2000; **19**:3005–3016.
38. Keiding N, Klein JP, Horowitz MM. Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine* 2001; **20**:1871–1885.
39. Klein JP, Shu YY. Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research* 2002; **11**:117–139.
40. Commenges D. Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research* 2002; **11**:167–182.
41. Aalen OO. A linear-regression model for the analysis of life times. *Statistics in Medicine* 1989; **8**:907–925.
42. Lin DY, Ying ZL. Semiparametric analysis of the additive risk model. *Biometrika* 1994; **81**:61–71.
43. Shu Y, Klein JP. Additive hazards Markov regression models illustrated with bone marrow transplant data. *Biometrika* 2005; **92**:283–301.
44. Norris JR. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press: Cambridge, 1997.
45. Lagakos SW, Sommer CJ, Zelen M. Semi-Markov models for partially censored data. *Biometrika* 1978; **65**: 311–317.
46. Gill RD. Nonparametric-estimation based on censored observations of a markov renewal process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 1980; **53**:97–116.
47. Prentice RL, Williams BJ, Peterson AV. On the regression-analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.
48. Dabrowska DM, Sun GW, Horowitz MM. Cox regression in a Markov renewal model: an application to the analysis of bone-marrow transplant data. *Journal of the American Statistical Association* 1994; **89**:867–877.
49. Farewell VT, Cox DR. A note on multiple time scales in life testing. *Applied Statistics* 1979; **28**:73–75.
50. Arjas E, Eerola M. On predictive causality in longitudinal studies. *Journal of Statistical Planning and Inference* 1993; **34**:361–384.
51. Aalen OO, Johansen S. Empirical transition matrix for nonhomogeneous Markov-chains based on censored observations. *Scandinavian Journal of Statistics* 1978; **5**:141–150.
52. Wangler M, Beyersmann J, Schumacher M. Changelos: an R-package for change in length of hospital stay based on the Aalen–Johansen estimator. *R News* 2006; **6**:31–35.

DOI: 10.1002/sim