Special Issue Paper

Statistics in Medicine

Received 3 December 2012, Accepted 27 August 2013 Published online 12 September 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5976

Multiple time scales in multi-state models

Simona Iacobelli[a,b*†] **and Bendix Carstensen**[c,d]

In multi-state models, it has been the tradition to model all transition intensities on one time scale, usually the time since entry into the study ('clock-forward' approach). The effect of time since an intermediate event has been accommodated either by changing the time scale to time since entry to the new state ('clock-back' approach) or by including the time at entry to the new state as a covariate. In this paper, we argue that the choice of time scale for the various transitions in a multi-state model should be dealt with as an empirical question, as also the question of whether a single time scale is sufficient. We illustrate that these questions are best addressed by using parametric models for the transition rates, as opposed to the traditional Cox-model-based approaches. Specific advantages are that dependence of failure rates on multiple time scales can be made explicit and described in informative graphical displays.

Using a single common time scale for all transitions greatly facilitates computations of probabilities of being in a particular state at a given time, because the machinery from the theory of Markov chains can be applied. However, a realistic model for transition rates is preferable, especially when the focus is not on prediction of final outcomes from start but on the analysis of instantaneous risk or on dynamic prediction.

We illustrate the various approaches using a data set from stem cell transplant in leukemia and provide supplementary online material in R.

Keywords: multi-state; time scales; Poisson model; time-dependent covariate; flexible parametric models

1. Introduction

Time-to-event histories with intermediate events are often analyzed in the framework of multi-state models, the simplest example being the illness-death model: one starting state ('well'), one intermediate state ('ill'), and one absorbing state ('dead'). The entry into the 'illness' state is supposed to change the hazard of occurrence of 'death'. In general, the illness-death model represents the occurrence of an event representing 'failure' as entry to the absorbing state and another event as entry in an intermediate state; and normally, it is assumed that it is not possible to go back to the initial state from the intermediate state.

A common, traditional approach, dating back to the 1970s in a famous (re)analysis of Stanford Heart Transplant study [1], is to model the hazard of occurrence of the failure event by a Cox regression, including the occurrence of the intermediate event as a time-dependent covariate, switching from 0 to value 1 at the time of occurrence. More recently, several papers indicated how to estimate all transition hazards and obtain transition probabilities in a multi-state model [2]. The most common approaches are by non-parametric (Nelson–Aalen method) or semi-parametric (Cox regression model) estimation [3]. A different approach based on pseudo-values avoids the estimation of transition hazards [4].

Using the non-parametric or semi-parametric estimator of the transition hazard, only one time scale is used. A time scale is basically a time-varying covariate that evolves deterministically. In some settings, the concept of time scale is a rather broad one, including any time-varying covariate or function of time; this approach is quite common in engineering [5, 6], but it was seen also in epidemiology [7] and evo-

[a] *Centro Interdipartimentale di Biostatistica e Bioinformatica, Università Tor Vergata, Rome, Italy*
[b] *Chronic Malignancies Working Party of the European Group for Blood and Marrow Transplantation, Leiden, the Netherlands*
[c] *Steno Diabetes Center, Gentofte, Denmark*
[d] *Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark*
**Correspondence to: Simona Iacobelli, Centro Interdipartimentale di Biostatistica e Bioinformatica, Università Tor Vergata, Rome, Italy.*
[†] *E-mail: simona.iacobelli@ebmt.org*

lutionary biology [8]. In this paper, we use only time scales that measure chronological time spent since a certain origin and advance at the same rate as this; therefore, time scales only differ in their origin. In clinical studies, it is natural to consider as time origin the moment when the individual starts to be subject to the risk of failure—typically the diagnosis of the disease or the entry into a trial. Usually, only one major relevant time scale is considered and chosen on the basis of *a priori* clinical knowledge; major criteria indicated in the literature for the choice of a time scale are in fact that the timescale is relevant for the context and such that the effects of the covariates of interest are interpretable [7, 9]; alternatively to the choice of one among several time scales of interest, approaches based on combining time scales into one have been proposed in the literature [7, 10]. Applying the standard methods of survival analysis, the unique time axis is used to define the risk sets, aligning individuals so that, given covariates, at any time they are homogeneous with respect to the risk of failure.

Multiple time scales originating from different origins (diagnosis, treatment, and onset of secondary disease) can, however, be relevant in clinical studies, where time scales such as current age (time since birth) or current calendar time (time since AD) are of lesser interest—as it is the case for the risk of death in the chronic myeloid leukemia (CML) example we use later. On the other hand, current age and current calendar time are usually the most relevant time scales in epidemiologic settings, where the follow-up is very extensive and the start of the (observational) study usually does not coincide with any event related to the onset of exposure or to any change of the risk of failure [11–13]. Traditional methods of demography and epidemiology describe the occurrence of events along multiple time axes [14–16].

In a multi-state analysis, one issue in the modeling process is the choice of the time scales. When intermediate states are considered, several time scales may become relevant for describing the risk of event; in an illness-death model, the risk may depend on time since start ('clock-forward' approach) as well as on time since entry to the intermediate state ('clock-back' or 'clock-reset' approach) [17]. Additionally, it can depend on the time-to-entry in the intermediate state; this variable is not a time scale but a covariate with a fixed value for all who enter the intermediate state. When using a Cox model with a time-dependent binary covariate for the occurrence of the intermediate event, it means implicitly that we assume that the baseline risk of failure depends only on time since start, with the same functional form for its effect before/without and after the intermediate event; in other words, because of the proportional hazards (PH) assumption, the intermediate event has only a multiplicative effect on the baseline hazard. It should be noted that a major motivation for having only one time scale in a multi-state model is that this will keep the underlying model as a Markovian model and thereby facilitate computations of transition probabilities, using the standard probability theory [3].

In the remainder of this paper, we shall argue that testing of the proportionality is not sufficient but that explicit modeling of (possible) non-proportionality is to be preferred. This also places the choice of one or more time scales as descriptors of the events rates as an empirical question. We show how the entire machinery is easily embedded in standard theory and implemented using the standard tools for generalized linear models. The case study we propose is useful to see how insight can be gained by using a modeling approach suitable to deal with multiple time scales as compared with the more traditional Cox PH regression.

2. The CML data

The series of data in the example analysis consists of 10,732 disease histories of patients with CML who received an allogeneic haematopoietic stem cell transplantation between 1980 and 2003 and were reported to the registry of the European Group for Blood and Marrow Transplantation. Death was observed for 4759 patients during a total follow-up of 36,611 person-years (overall mortality rate 13% per year). The most relevant intermediate event during the disease history is the occurrence of relapse, which was observed in 2246 patients during 30,504 person-years follow-up (relapse rate 7.1% per year, mortality rates 17.7% per year post relapse; Figure 1). Time to relapse varied from less than 1 month to 17 years after transplantation, with a median equal to 12 months. The main initial objectives of the study were to assess the impact of relapse occurrence on mortality and, as a secondary question, to illustrate its evolution during the entire period when stem cell transplantation was the most effective treatment for CML.

The natural time scale as in any study on transplanted patients is time since transplant; additionally, the occurrence of relapse defines another relevant time scale for the subsequent risk of death, time since relapse. Risk factors for mortality after transplantation in CML identified in the literature

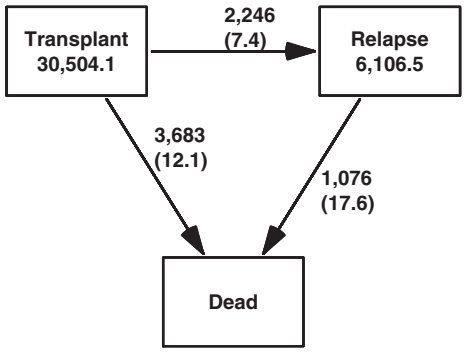

Figure 1. Distribution of follow-up time and events between states. The numbers in the boxes are total follow-up times (person-years), and the numbers on the arrows are number of events, with rates per 100 person-years. Initial population: $n = 10,732$.

are advanced disease stage at transplant (blast crisis/accelerated phase/chronic phase), type of donor (matched unrelated donor/human leukocyte antigen-identical sibling donor) and prolonged time interval between diagnosis and transplant (>1 year/⩽1 year). These three factors were included in all models throughout this paper, with reference group being the patients transplanted in chronic phase from a human leukocyte antigen-identical sibling donor less than or equal to 1 year after diagnosis (this is a low-risk group, and the largest, $n = 4200$ patients). Additionally, we included calendar year and age both measured as time-fixed covariates at transplant. A possible extension of the current analysis is to investigate the role of current calendar time as a third time scale of interest, and its interaction with relapse occurrence, to answer to the second original research question. This is, however, not pursued here.

Time from transplant to relapse is a potential explicative factor for mortality after relapse, as described in stem cell transplant literature. However, as we will show in Section 3, the interpretation of its role is linked to the choice of the relevant time scale(s), and benefits from a broad investigation carried out with the approach that we propose.

We do not consider other potential relevant factors for mortality after relapse such as type of relapse or type of treatment received; thus, the analyses are proposed with no intent of delivering messages to be considered for clinical decision making. Also, note that if we want to include the effects on the risk of failure of characteristics originating after relapse, this can be carried out by introducing time-dependent covariates in the models for the transition hazard, but in order to compute transition probabilities from the entry state, it would require a model for the distribution of such factors. In other words, different characteristics of/after relapse should be represented in terms of additional intermediate states, and we should be able to model the entry into such states.

An overview of the patients' flow is given in Figure 1. A full account of all analyses with R codes is given in the online supplementary material[‡] http://bendixcarstensen.com/AdvCoh/MS/BackgroundAnalyses.pdf. It includes a table of patients' characteristics and a display of (a selection of) life histories from the CML data in Lexis diagrams [18], along different time scales.

3. Models of the mortality rates after relapse

3.1. Notation

We denote with t the time since transplant, with d the time since relapse (duration), and $r = t - d$ is then time from transplant to relapse, or time to entry into the intermediate state. Note that r is a covariate that is only defined after relapse, but it is not a time scale, as it is constant during the follow-up in the 'relapse' state.

The symbol x will be used to indicate a vector of fixed covariates. In our models for CML data, we model the effect of calendar year and age at transplant by use of restricted cubic splines with 6 knots (five parameters for each spline); thus, the dimension of x is 14 (the covariates mentioned in Section 2—excluding time from transplant to relapse—require four parameters). In all succeeding formulas, all quantities are scalars except x and the corresponding vector of regression coefficients γ.

[‡]*Supporting information may be found in the online version of this article.*

The letter μ will be used for the transition hazards into the death state, and λ will be used for the transition into the intermediate (relapse) state. For the multi-state model with states and transitions illustrated in Figure 1, $X(t)$ will indicate the status occupied at time t, which could be transplant (T), relapse (R), or death (D).

3.2. *Cox models with one time scale: clock-back and clock-forward approaches*

One 'piece' of the multi-state model for the disease history of CML is the transition from relapse to death. In studies investigating the outcomes after relapse in CML and thus restricted to relapsed patients, the usual approach is to model the risk of death along the scale of time since relapse. This is the clock-back approach, where the survival is computed since entry in the current state. There is a clear clinical motivation for this choice: It is expected on biological grounds that the patient undergoes a higher risk of mortality in the early phase after relapse and that later on the risk decreases because of effective salvage treatment (and death of the most ill patients); thus, variation in mortality is anticipated to be largest on this time scale. There is also a practical motivation: the methods of analysis such as the Kaplan–Meier survival curve are simpler when all observations start at time 0.

The effect of the total duration of the disease since transplantation is usually taken into account by including (a function of) the time interval between transplant and relapse as a (time-fixed) covariate in the model. Two retrospective studies conducted on European Group for Blood and Marrow Transplantation data showed that longer times from transplant to (clinical) relapse were associated with higher survival probability [19, 20]. This protective effect could be interpreted considering longer time to relapse as a marker of less aggressive disease, or better eradication of the leukemia.

In the initial analysis of CML data, we can thus use the following model for analysis of mortality rates among relapsed patients:

$$\mu(d) = \mu_0(d)\exp(\beta r + \gamma x) \tag{1}$$

The estimate for the hazard ratio (HR) per year of time from transplant to relapse ($\exp(\beta)$) is 0.80, so we see that the mortality decreases by the time from transplant to relapse among those who relapse, consistently with the existing medical literature.

An alternative to the clock-back approach is the clock-forward approach, which uses delayed entry (left-truncation), that is, takes as outcome variable the triple given by time of entry in the relapse state, survival time since transplant, and survival status indicator, and constructs the risk set at time t_j as containing all individuals with entry time r before t_j and the exit time after t_j (because the relapse times vary from patient to patient, the risk sets and thus the partial likelihood are different from the clock-back approach). It is very uncommon in a study restricted to relapsed patients to use time since transplant as time scale; on the other hand, the clock-forward approach is what is implicitly assumed when analyzing the entire disease history from transplantation, including also patients who do not relapse and for those who do using a time-dependent covariate, which switches from 0 to 1 at time of relapse (see model 4).

The Cox clock-forward model for mortality after relapse for relapsed patients in the CML study is as follows:

$$\mu(t) = \mu_0(t)\exp(\beta r + \gamma x) \tag{2}$$

The estimate for the HR per year of time from transplant to relapse using the clock-forward approach is 1.18, so we see that when viewed on this time scale, the mortality increases by the time from transplant to relapse among those who relapse.

This result, which may appear contradictory with respect to the clock-back approach and counterintuitive for the lack of a biological explanation, is because the hazard decreases sharply with time since relapse: So for a fixed time since transplant, a longer time to relapse means shorter time since relapse and thus higher risk of death. The difficulty of interpretation emerges from the un-natural choice of the time scale in a model that allows to use only one time scale for the baseline hazard. The graphical representations given in Figures 3 and 4, derived from the alternative approach that we will suggest in Section 4, provide a simple graphical explanation of this phenomenon.

The issues with effects of time to/since relapse are particular to this data set and merely serves as an illustration that counterintuitive results can emerge. The general point is, however, that statements of effect sizes in a Cox model are meaningless unless taken in conjunction with the choice of the underlying time scale. The solution to this time-scale problem is, however, not speculative; in our example, for

patients with a relapse, it is an empirical question whether the mortality depends on time since transplant, time since relapse, time at relapse, or all three. We shall look into this in the next sections.

3.3. Cox models with multiple time scales

The risk of death after relapse may depend on both t (time since transplant) and d (time since relapse). As we have seen respectively in models (1) and (2), both the clock-back (using d) and clock-forward (using t) approaches to Cox regression try to overcome the limitation of using only one time scale for the baseline hazard by including dependence on the second time scale through the linear effect of time to relapse (i.e., the constant difference of the two time scales). It is easy to show that this is equivalent to control for the *linear* effect of the second time scale. For example, model (2) is as follows:

$$\begin{aligned}\mu(t) &= \mu_0(t)\exp(\beta r + \gamma x)\\ &= \mu_0(t)\exp(\beta(t-d)+\gamma x)\\ &= \mu_0(t)\exp(\beta t)\exp(-\beta d+\gamma x)\\ &= \tilde{\mu}_0(t)\exp\left(\tilde{\beta} d+\gamma x\right)\end{aligned}$$

In these expressions, what is relevant is that while r is a covariate assuming a constant value along the whole follow-up after relapse, in the latter expression, d is a time scale, namely the current time since relapse, that is, a variable whose value is continuously updated during the follow-up. In practical terms, the latter model could be fit by splitting the follow-up of each person into small time intervals, computing the current time since relapse (and possibly the value of other time scales too) for each interval and including it in the model as a covariate (with linear effect). Illustration of the split data is given in the online supplementary material, based on the Lexis machinery of the Epi package [21].

Thus, models (1) and (2) are special cases of Cox models with two time scales, t and d, where one time scale is chosen as the reference time scale for the baseline hazard and the second time scale is included as a covariate with linear effect on the log-hazard scale. To make the model more general and sufficiently flexible to investigate the role of the second time scale, it is necessary to include a non-linear effect of it using a suitable smooth parametric function (splines, for example).

This approach to multiple time scales, however, seems rather artificial because the effect of one time scale on the hazard is modeled non-parametrically (or, which is equivalent, with one parameter per event time [22]) and actually treated as a nuisance object, while the effects of other timescales are modeled with a smooth parametric function. This is an intrinsic limitation of the Cox regression, because the model is conceived to disregard the role of timing by assuming proportionality between hazards and using the partial likelihood for the estimation of the regression coefficients. If multiple time scales are to be investigated, it seems more natural to model the effect of all of them explicitly and in the same way.

3.4. Poisson models with multiple time scales

The following two-way PH model [23] includes the dependence on two time scales in a symmetric and simple manner:

$$\mu(t,d) = \mu_0(t)\phi_0(d)\exp(\gamma x) \tag{3}$$

The estimation of this model can be carried out by using a split data set as input to Poisson regression. The data split along time is necessary to compute the current value of each time scale for each interval, as seen in the previous section. The effects of the time scales t and d can be modeled parametrically through a smooth function, thus assuming a flexible shape of the baseline hazards, the argument of the function being the value of that time scale at the start of each small time interval.

Notice that this approach might be seen as a natural generalization of the Cox-based approaches illustrated in the previous sections. It is in fact well known that in a PH model with constant baseline hazard in each of a sequence of small intervals, the contribution to the likelihood of the follow-up of one person through a number of intervals is equivalent to that of independent Poisson-distributed observations; thus, instead of using the Cox partial likelihood, it is possible to obtain the same estimates of the regression coefficients by applying a Poisson regression, including the baseline hazard as a parametric function and the log of the length of the time interval as an offset [24]. Table I shows that the estimates of the regression parameters in Cox and in Poisson are identical for all practical purposes, even for moderately large intervals and a fairly simple parametrization of the baseline hazard.

Table I. Estimates of covariates effects on mortality after relapse: (a) model 1, (b) model 2, and (c) model 3.

		Poisson			Cox		
		exp(coeff.)	95% CI lower	95% CI upper	exp(coeff.)	95% CI lower	95% CI upper
(a) Clock-back (time since relapse)	diag-trx > 1 year	1.04	0.92	1.19	1.05	0.92	1.19
	stage AcclPhase	2.34	2.01	2.71	2.33	2.01	2.7
	stage BlastCrisis	5.75	4.84	6.83	5.6	4.72	6.66
	donor Other	1.25	1.06	1.48	1.24	1.05	1.46
	time to relapse	0.80	0.76	0.84	0.84	0.80	0.88
(b) Clock-forward (time since transplant)	diag-trx > 1 year	1.05	0.92	1.19	1.04	0.92	1.19
	stage AcclPhase	2.14	1.85	2.49	2.13	1.83	2.47
	stage BlastCrisis	4.34	3.62	5.20	4.27	3.55	5.12
	donor Other	1.18	1.00	1.40	1.18	1.00	1.39
	time to relapse	1.18	1.12	1.24	1.18	1.12	1.25
(c) 2-Way PH model (both time scales)	diag-trx > 1 year	1.05	0.92	1.19			
	stage AcclPhase	2.15	1.85	2.49			
	stage BlastCrisis	4.35	3.63	5.22			
	donor Other	1.19	1.01	1.40			

The estimation approach for model (3) relies on an appropriate choice of the time split [23]. As the Poisson model assumes that the hazard rate is constant in each time interval, the follow-up data have to be split in small intervals, which increases the computational burden, although it is feasible with current technology. The split along the time scale should be very fine in the region where we suspect the largest variation in the event rate (in the CML data, at the beginning of the follow-up) and could be coarser elsewhere. For the models on CML data, we split time in 2-month intervals for the first 3 years, 6-month intervals the next 2 years, and 1-year intervals subsequently; and for modeling, we used restricted cubic splines for both time since transplant and time since relapse with five parameters. The work by Efron [23] remarks that this multiplicative 'two-way PH model' can be traced back to Lexis in the 1870s and provides some insight in theoretical issues. The choice between Cox and Poisson models was discussed for example in [2].

The advantages of this approach is that it is possible to accommodate several time scales simultaneously, that all time scales are treated symmetrically, and that each time scale has an impact on the rate in terms of a smooth function, which facilitates the graphical representation of the hazard rate itself; the relevance of having graphical displays will be illustrated in Section 4.1. It is also useful for testing the effects of time scales, which is easily carried out as comparisons of nested models by likelihood-ratio test. All tools for the data split are available from the Epi package in R [21].

4. Models of the mortality rates for the entire follow-up

The traditional approach to modeling the mortality along the whole disease history is to use the following Cox model with time-dependent covariates:

$$\mu(t) = \begin{cases} \mu_0(t)\exp(\gamma x) & \text{before relapse,} \quad t < r \\ \mu_0(t)\exp(\gamma x + \beta_{\text{rel}} + \beta r) & \text{after relapse,} \quad t > r \end{cases} \tag{4}$$

where the baseline hazard varies along only one time scale, the time since transplant; the risk of failure after relapse varies along the same time scale (clock-forward approach) and depends also linearly on time to relapse. We have seen in Section 3.3 that this adjustment actually corresponds to a particular model based on the dependence of the hazard of death of relapsed patients on both time since transplant and time since relapse, but that within the framework of Cox regression there is only a limited possibility of investigating the role of both time scales in a natural way.

Instead, following the aforementioned approach of maximizing the full likelihood of a parametric Poisson model with flexible baseline hazard function(s), we can include dependence on all time effects

thought to be relevant for the risk of death after relapse. The model we will use for the CML data is as follows:

$$\mu(t,d) = \begin{cases} \mu_0(t)\exp(\gamma x) & \text{before relapse,} \quad t < r \\ \mu_0(t)\phi_0(d)\exp(\gamma x + \beta_{\text{rel}} + \eta(r)) & \text{after relapse,} \quad t > r \end{cases} \tag{5}$$

This model includes the effects of both time scales t and d and of time to relapse r (the latter by a non-linear function $\eta(\cdot)$, fitted again using a cubic spline with 6 knots). This might at a first glance seem unfeasible, because the latter is always equal to the difference of the two time scales $r = t - d$, thus suggesting the classical problem of identifiability of age–period–cohort models [16]. However, when we model the risk of death using the whole disease history, we can estimate the effect of all three variables t, d, and r, because the linear relation $r = t - d$ does not hold anymore, as the time to relapse (r) is formally undefined prior to relapse (and set equal to 0 for practical purposes), as is time since relapse (d). The intuition behind this mathematical structure that allows the identification of the three effects is that we can estimate the effect of time since transplant (t) largely based on persons without relapse, and the additional effects of r and d are estimated only from those who relapse. In other terms, in the analysis of mortality of relapsed patients, we introduce 'external' information (from disease history prior to relapse), which imposes a certain structure to the effect of one time scale and thus allows the estimation of the effects of the other two timing variables (for references, see [14]).

A model like (5) could easily be extended to include further time scales. For example, for the CML data, we might investigate the role of current calendar time, representing changes in the transplant procedures and in detection and treatment of relapse. In studies of chronic diseases with longer duration and follow-up, also, current age could be a relevant scale for the evaluation of the risk of failure.

4.1. Reporting the effects of multiple time scales

Graphical representations are necessary to report the effects of multiple time scales and, at the same time, are very informative for the investigators. One type of representation of model (5) fitted to the CML data is given in Figure 2. It shows a 'decomposition' of the effects of the time variables, which act multiplicatively on the total mortality rate. The first panel shows the effect of the time since transplant, by giving the variation of the absolute mortality rate along this time axis, for a patient with a fixed set of covariates and no relapse. The second panel shows the effect of the second time scale, time since

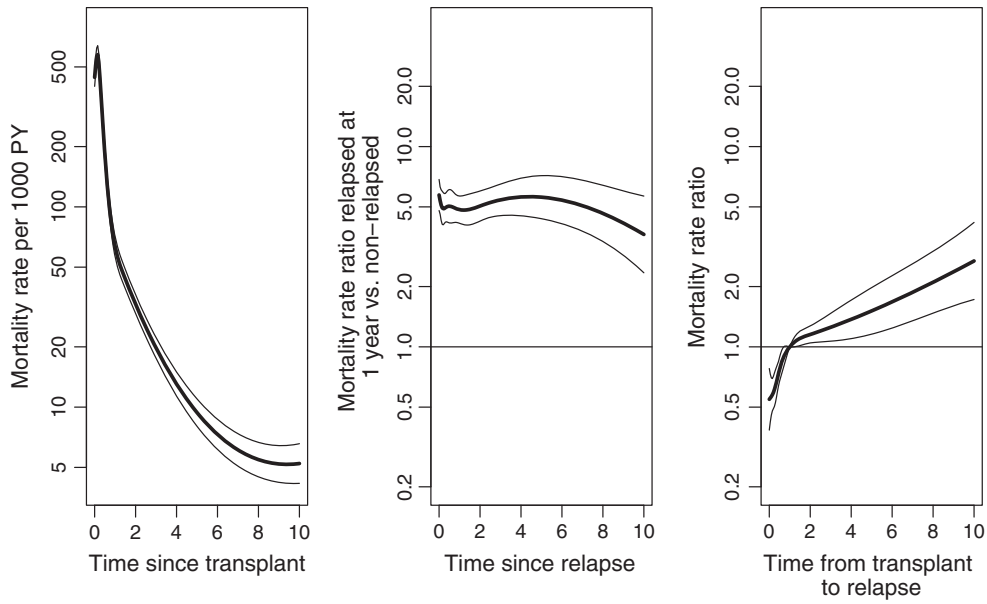

Figure 2. Model 5: Estimated effects of timing for the CML data, with 95% confidence intervals. Patients in chronic phase, transplanted within 1 year after diagnosis and with a sibling donor. Panel 1: Absolute rates for no-relapse patients, effect of time since transplant. Panel 2: Rate ratio for patients relapsed 1 year after transplant versus patients without relapse, effect of time since relapse. Panel 3: Effect of time to relapse (rate ratio), relative to patients relapsed at 1 year after transplant.

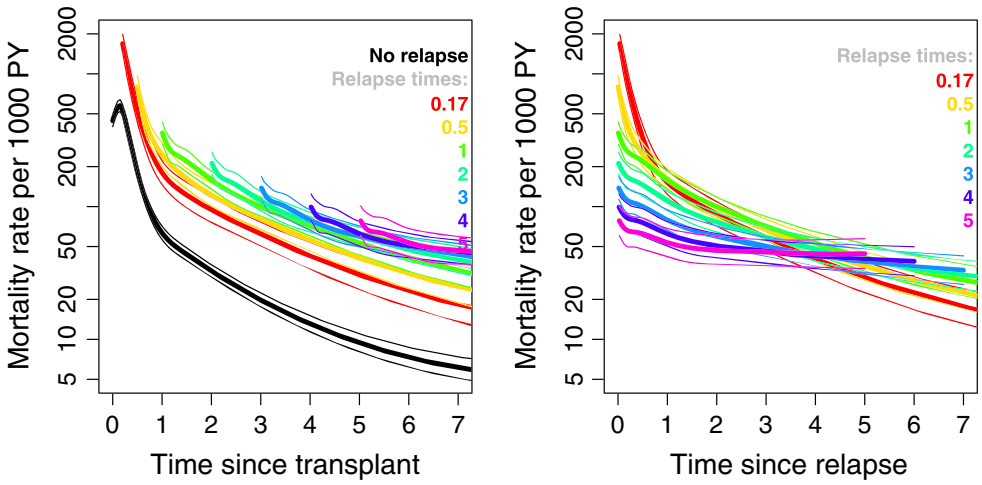

Figure 3. Model 5: Estimated mortality rates for patients without relapse and for patients with different relapse times. Mortality rates for relapsed patients are plotted on two different time scales: on the left, versus time since transplant, and on the right, versus time since relapse. Patients are in chronic phase, transplanted before 1 year after diagnosis and with a sibling donor.

relapse, in relative terms (rate ratio) relative to patients without relapse; the comparison is for the same time since transplant t and for the same fixed pattern of covariates. As the model also includes the effect of time to relapse, for this plot, we assumed that relapse occurred at 1 year since transplant ($r = 1$). The third panel shows the impact of a different time to relapse r, again in terms of rate ratio (in other terms, the baseline is the patient with $r = 1$ considered in the second panel). All graphs are reported on the log-scale on the y-axis and, for example, a ratio of 2 is represented by the same vertical distance in all panels. This is carried out to facilitate the 'reconstruction' of the rate by adding the effects of its components.

Figure 2 is informative in particular for model specification purposes, as it shows the relevant impact of time since transplant t (first panel) and, among relapsed patients compared at the same t, the limited effect of time since relapse, illustrated by a substantially constant rate ratio in the second panel. Finally, among relapsed patients, the effect of time until relapse is more pronounced (third panel), with higher relative risk associated with longer time to relapse. This latter finding is in line with the result ($HR = 1.18$) obtained in the clock-forward Cox model (2) using time since transplant as baseline time scale, since as just remarked in the present model the contribution of the second time scale is negligible, which makes the model quite similar to model (2).

On the other hand, Figure 2 is of little use in assessing the general variation in time of the risk of death for all patients, since, as noted earlier for the relapsed patients, the marginal effects have to be added up, a difficult exercise when it has to be performed in your head. This is particularly important for time scales, because time scales vary together, and the usual regression interpretation as 'effect for all other variables fixed' does not apply. Hence, in order to show how mortality varies with time since transplant for patients without relapse and patients that relapse at different times and by time since relapse for those who relapse, we computed the estimated mortality rates, both by time since transplant and by time since relapse, as shown in Figure 3.

Figure 3 explains the apparently contradictory results from the clock-forward and clock-back approaches seen in Section 3.2 regarding the role of time to relapse. In the clock-forward model, we refer to the left panel where successively later relapses result in higher mortality for a given time since transplant. In the clock-back model, we refer to the right panel where successively later relapses result in lower mortality for a given time since relapse. This trend is the result of the combination of the effects displayed in the first and third panels of Figure 2. When considering a fixed time since relapse, longer time to relapse implies longer time since transplant, and because the latter shows a curve decreasing more steeply than the increase of the rate ratio for time to relapse, the overall effect is a reduction of the total rate. Note that this is not a general phenomenon but a particular feature of the data set at hand. In the next section, we will see that this phenomenon is visible also when making a different choice of which time effects to include in the model (Figure 3).

4.2. *Testing time-scale effects*

All effects in a generalized linear model, including the effects of time scales, can be assessed by the likelihood-ratio test to compare nested models, specifically with and without the effect to be assessed. To assess the presence of non-linear effect of a timing variable, it is sufficient to compare the model including the spline transform to the model with the linear effect only. Our graphical analysis suggests that the effect of time since relapse could be removed from the model for parsimony. The results of the likelihood-ratio tests for the various Poisson models are shown in Table II. It basically shows that there is no overwhelming evidence for any effect of time since relapse, only of time to relapse, basically confirming that the middle curve in Figure 2 is not different from a horizontal line. Hence, the relevant description of data would be using a model where the mortality depends on time since transplant and for the relapsed patients additionally on time to relapse. Our final model is thus as follows:

Table II. Likelihood-ratio tests for model choice with respect to timing effects.

Effect	d.f.	χ^2	P
Non-linear effect of time to relapse vs. linear	4	21.09	<0.001
Linear effect of time to relapse vs. no effect	1	57.44	<0.001
Non-linear effect of time since relapse vs. linear	4	9.55	0.049
Linear effect of time since relapse vs. no effect	1	1.04	0.307

Subsequently removing effects from model 5.

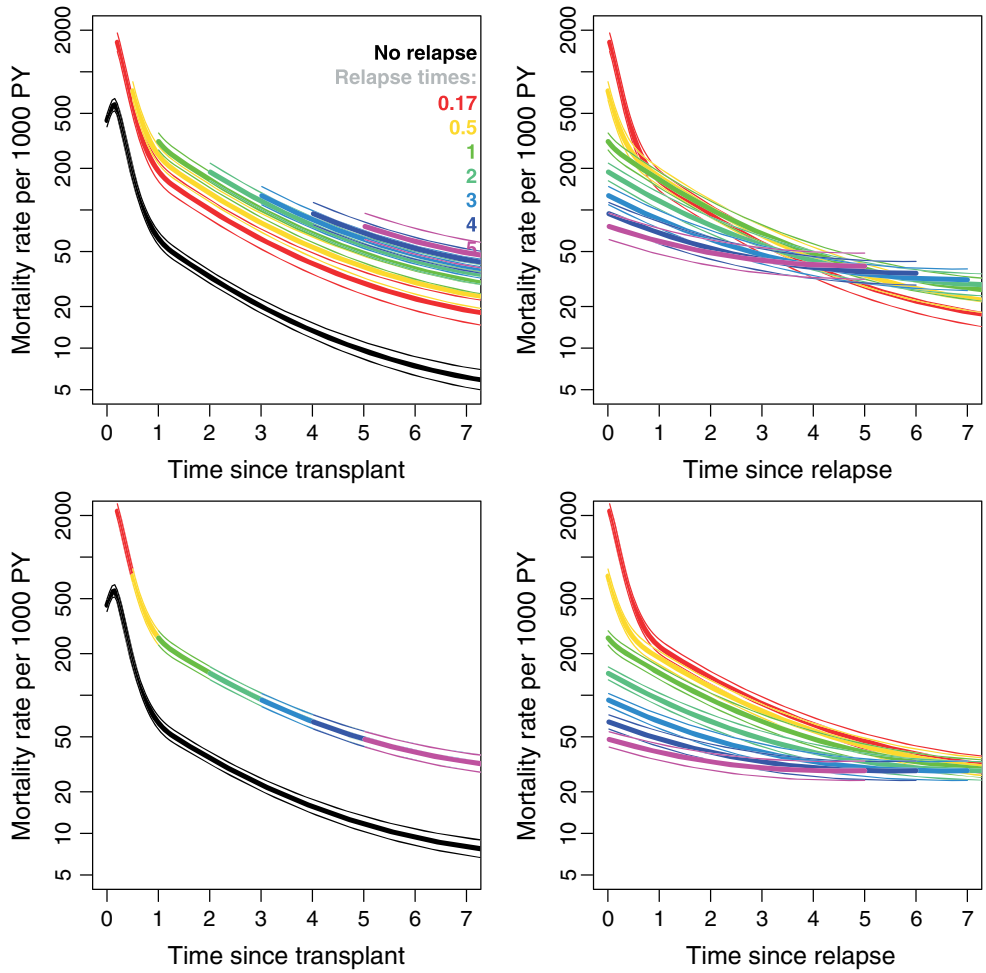

Figure 4. Model 6 compared with a Markovian model: Rates. Top panels: Estimated mortality rates for patients without relapse and for patients with different relapse times from a model with baseline hazard on time since transplant and non-linear effect of time to relapse (model 6). Bottom panels: corresponding quantities estimated from a Markovian model with rate of mortality after relapse depending only on time since transplant.

$$\mu(t,d) = \begin{cases} \mu_0(t)\exp(\gamma x) & \text{before relapse,} \quad t < r \\ \mu_0(t)\exp(\gamma x + \beta_{\text{rel}} + \eta(r)) & \text{after relapse,} \quad t > r \end{cases} \tag{6}$$

The estimated mortality rates under this final model are shown in Figure 4 (top panels).

5. Transition probabilities

To complete the multi-state model describing the course of CML after transplantation, we need a model for the relapse rate $\lambda(t)$. We estimated the relapse rate using a smooth parametric function of time since transplant and the same fixed covariates as in the models for the mortality rate.

Denoting with μ_{TD} and μ_{RD} the mortality rates (transition to D) with starting state the initial one (T) and the intermediate one (R), respectively, standard theory gives the following expressions of the probabilities in terms of the transition rates:

$$\begin{aligned} P_{TT}(s,t) &= P(X(t) = T | X(s) = T) = \exp\left(-\int_s^t (\lambda(u) + \mu_{TD}(u))\,\mathrm{d}u\right) \\ P_{RR}(s,t;r) &= P(X(t) = R | X(s) = R, T_1 = r) = \exp\left(-\int_s^t \mu_{RD}(u)\mathrm{d}u\right) \\ P_{TR}(s,t) &= P(X(t) = R | X(s) = T) = \int_s^t P_{TT}(s,u^-)\lambda(u)P_{RR}(u^+,t;r)\mathrm{d}u \\ P_{TD}(s,t) &= P(X(t) = D | X(s) = T) \\ &= \int_s^t P_{TT}(s,u^-)\mu_{TD}(u)\mathrm{d}u + \int_s^t P_{TT}(s,u^-)\lambda(u)\left(1 - P_{RR}(u^+,t;r)\right)\mathrm{d}u \end{aligned} \tag{7}$$

Analogous formulas can be written for discrete time, which can be used for practical calculations. More general formulas for other models can be found for example in [2, 3].

The two terms of the sum returning P_{TD} are two relevant transition probabilities, represented by two separate curves in Figure 5, left panel. One is the probability of direct transition from state T to state D, that is, the probability of death without relapse. The second is the probability of passage from T to R and then from state R to state D, that is, the probability of dying having experienced relapse.

In the Markovian case with only time since transplant as time scale, the calculation of these quantities and their standard errors is relatively simple (see, for example, [3] or [25]), which also illustrates the use

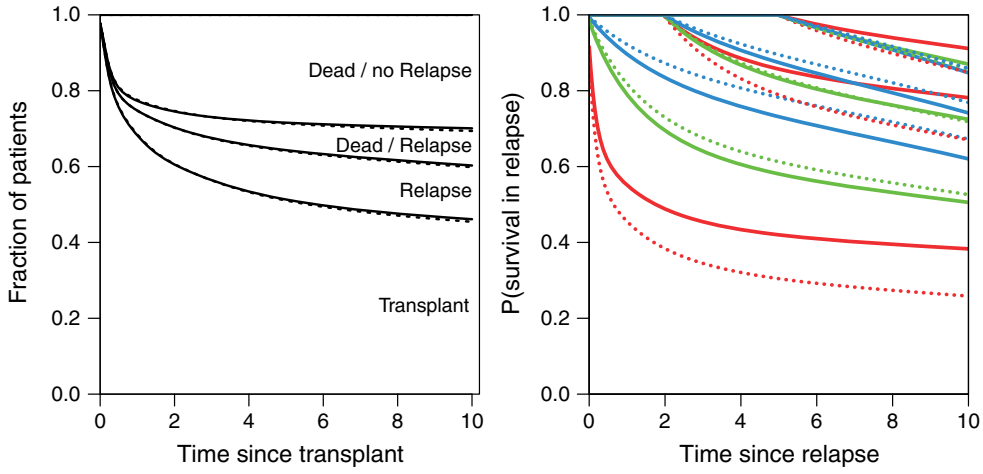

Figure 5. Model 6 compared with a Markovian model on time since transplant. Left panel: Transition probabilities from initial state (transplant) and prediction time equal to day of transplant. Model 6 (dotted black) compared with a Markovian model (solid red). Right panel: Transition probabilities from the intermediate state (relapse) and prediction time equal to 0, 2, and 5 years after relapse, into the same state at later time. They are the probabilities of survival in 'relapse' state. Model 6 (solid) compared with a Markovian model with only time since transplant (dotted). Time to relapse is equal to 2 months (red), 1 year (green), and 3 years (blue) since transplant.

of the R package mstate for multi-state models with non- and semi-parametric approaches. In our non-Markovian setting with death after relapse depending on time to relapse, there is no simple machinery to compute the transition probabilities, but the calculations based on Equation (7) are facilitated by having fully parametric models. Rates have a closed-form expression, so they can be computed at any desired point of the timescales. Furthermore, it is possible to assess the precision of the estimated transition probabilities by simulation (sometimes called parametric bootstrap), basically taking a random point from the 'posterior' distribution of the estimates (that is, the multivariate normal with mean equal to the estimate and variance equal to the estimated variance–covariance from the Poisson models), computing the transition probabilities as a function of time, and repeating the procedure M times. From the sample of M sets of estimated transition probabilities, we can then construct 95% confidence limits by taking the 2.5^{th} and 97.5^{th} percentiles.

Figure 5 compares on the left panel the estimated transition probabilities from the initial state at time 0 under two different models, the fully parametric model we selected (model 6) and a Markovian model where we removed the effect of time to relapse on the mortality rate after relapse. Interestingly, despite the latter model appeared to be an inadequate description of the mortality rates, the transition probabilities appear to be robust with respect to the mis-specification. However, other quantities are relevant in the analysis, in particular those related to dynamic prediction, that is, prediction incorporating information on events occurred after the entry into the study. When looking at the estimated transition probabilities from the intermediate state (Figure 5, right panel), the effects of mis-specification are pretty dramatic. As it can be seen in Figure 4, the Markovian model returns a mortality rate that is higher with very short time to relapse and lower with long time to relapse, with respect to model 6. As a consequence, the probability of survival after relapse in the Markovian setting is underestimated for time to relapse equal to 2 months and overestimated for longer time to relapse.

6. Discussion

The analysis of occurrence rates in multi-state models should naturally consider several time scales, possibly the time since entry into the initial state (when this is not relevant like in observational studies where the entry data are simply the date of inception of a register, natural time scales are calendar time or current age) and the time since entry into any transient state. As seen in the CML example we presented here, the difference between two scales, namely the time at entry into a given state, may be of interest, too. The usual approaches to multi-state models adopt, however, rates estimation methods (Nelson–Aalen estimator and Cox PH model) that consider only one time scale, and literature gives little suggestions on how to choose it. Markovian models, that is, clock-forward approach and no effect of entry times are preferred largely from a practical point of view, to facilitate the estimation of transition probabilities [3, 25].

The CML study is an example where neglecting the role of time to entry into an intermediate state, or modeling it by a rigid structure, is not satisfactory from neither the point of view of the interpretation nor for an adequate model fit. We also saw that the interpretation of simple regression parameters for these timing variables may be difficult without resort to a graphical representation of the mortality rates, because the actual shape of the baseline rates plays a central role. In order to be able to estimate and plot mortality rates, and in particular if these rates depend on more than one time scale, it is most convenient to base modeling on parametric models. We used a natural extension of a PH model that allows dependence on multiple time scales assuming a multiplicative structure (Efron two-way PH model), using Poisson regression for time-split data and smooth parametric functions to model dependence on time. It is in fact possible to accommodate more than one time scale in a Cox model, but they will not be treated algorithmically the same way, and the shape of the effect will be better investigated for the time scales other than the primary. The Poisson approach does not require additional computational burden for modeling than the Cox approach, apart from arranging the data and splitting the follow-up along the relevant time scales. This is easily carried out using the Lexis machinery in the Epi package for R in [21], which provides also tools useful for the estimation of rates. The approach is thus feasible also when the sample size and the number of events are more limited than in our study (model 5 has 15 parameters more than model 4, where the effect of time to relapse is linear). An additive model alternative to the Efron model was proposed in [26].

Our approach to the analysis of transition rates is easily extended to multi-state models with more states and transitions than an illness-death model, provided the adequacy of the number of transitions observed, and it is actually the most appropriate because it would be more difficult to select the relevant

time scales only on the basis of clinical knowledge. As we have seen, the selection can be carried out by means of simple likelihood-ratio test for nested models.

Of course when the transition hazards of a multi-state model are estimated from models with multiple time scales, there is an important limitation that the computation of transition probabilities becomes convoluted or even practically unfeasible in models with many states and transitions, because of the loss of the Markovian property. However, we have seen that predictions from intermediate states can be heavily affected by mis-specification of dependence on timing. Simulation approaches for the calculation of transition probabilities should thus be used for complex multi-state models. In the Epi package for R, this is carried out by the simLexis function.

The issue of investigation of the role and choice among multiple time scales is relevant also in the framework of standard survival models (multi-state models with only two states). One broad category is constituted by the studies where the observations can be considered left truncated, for example, when treatment starts some time after diagnosis, as in all studies on patients who undergo a haematopoietic stem cell transplantation, which, depending on the disease, might be given even after several years from diagnosis; other examples are observational studies where the diagnosis is carried out after a varying time since disease onset or since start of exposure. Another category is the one of clinical studies with very long follow-up, where calendar time and current age should always be taken into account.

To conclude, we recommend to investigate the dependence on multiple time variables using flexible parametric modeling instead of the usual Cox-based approach, especially if the focus is on the variation of the instantaneous risk in time and on dynamic prediction.

Acknowledgements

This work has originated from a study of the Chronic Malignancies Working Party of the European Group for Blood and Marrow Transplantation. The authors are grateful to Nicolaus Kroger on behalf of the CMWP and to Eduardo Olavarria on behalf of the Myeloproliferative Neoplasia Subcommittee for the use of data and for interesting discussions.

References

1. Crowley J, Hu M. Covariance analysis of heart transplant survival data. *Journal of the American Medical Association* 1977; **72**(357):27–36.
2. Andersen PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 2002; **11**:91–115.
3. Andersen PK, Perme MP. Inference for outcome probabilities in multi-state models. *Lifetime Data Analysis* 2008; **14**(4):405–431.
4. Andersen PK, Klein JP. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics* 2007; **34**:3–16.
5. Duchesne T, Lawless J. Alternative time scales and failure time models. *Lifetime Data Analysis* 2000; **6**:157–179.
6. Kordonsky KHB, Gertsbakh I. Multiple time scales the lifetime coefficient of variation: engineering applications. *Lifetime Data Analysis* 1997; **2**:139–156.
7. Oakes D. Multiple time scales in survival analysis. *Lifetime Data Analysis* 1995; **1**:7–18.
8. van Dooren TJM, Tully T, Ferrièr R. The analysis of reaction norms for age and size at maturity using maturation rate models. *Evolution* 2005; **59**:500–506.
9. Cook RJ, Lawless JF. *The Statistical Analysis of Recurrent Events*. Springer: Berlin, Heidelberg, New York, 2007.
10. Farewell VT, Cox DR. A note on multiple time scales in life testing. *Journal of the Royal Statistical Society* 1979; **28**(1):73–75. series C.
11. Thiébaut ACM, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine* 2004; **23**:3803–3820.
12. Korn EL, Graubard BL, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time scale. *American Journal of Epidemiology* 1997; **145**(1):72–80.
13. Pencina MJ, Larson MG, D'Agostino RB. Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* 2007; **26**:1343–1359.
14. Carstensen B. Age–period–cohort models for the Lexis diagram. *Statistics in Medicine* 2007; **26**:3018–3045.
15. Clayton D, Schifflers E. Models for temporal variation in cancer rates I: age–period and age–cohort models. *Statistics in Medicine* 1987; **6**:449–467.
16. Clayton D, Schifflers E. Models for temporal variation in cancer rates II: age–period–cohort models. *Statistics in Medicine* 1987; **6**:469–481.
17. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* 2007; **26**(11):2389–430.
18. Lexis W. *Einleitung in die Theorie der Bevölkerungsstatistik*. Karl J Trübner: Strassburg, 1875.

19. Arcese W, Goldman JM, D'Arcangelo E, Schattenberg A, Nardi A, Apperley JF, Frassoni F, Aversa F, Prentice HG, Ljungman P. Outcome for patients who relapse after allogeneic bone marrow transplantation for chronic myeloid leukemia. Chronic leukemia working party. European Bone Marrow Transplantation Group. *Blood* 1993; **82**:3211–3219.
20. Guglielmi C, Arcese W, Hermans J, Bacigalupo A, Bandini G, Bunjes D, Carreras E, Devergie A, Frassoni F, Goldman J, Gratwohl A, Kolb HJ, Iori AP, Niederwieser D, Prentice HG, de Witte T, Apperley J. Risk assessment in patients with Ph+ chronic myelogenous leukemia at first relapse after allogeneic stem cell transplant: an EBMT retrospective analysis. The chronic leukemia working party of the European Group for Blood and Marrow Transplantation. *Blood* 2000; **95**:3328–3334.
21. Plummer M, Carstensen B. Lexis: An R class for epidemiological studies with long-term follow-up. *Journal of Statistical Software* 2011; **38**(5):1–12.
22. Whitehead J. Fitting Cox's regression model to survival data using GLIM. *Applied Statistics* 1980; **29**(3):268–275.
23. Efron B. The two-way proportional hazards model. *Journal of the Royal Statistical Society* 2002; **64**(4):899–909. series B.
24. Efron B. Logistic regression, survival analysis and the Kaplan–Meier curve. *Journal of the American Medical Association* 1988; **83**(402):414–425.
25. de Wreede LC, Fiocco M, Putter H. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine* 2010; **99**(3):261–274.
26. Kauermann G, Khomski P. Additive two-way hazards model with varying coefficients. *Computational Statistics and Data Analysis* 2006; **51**:1944–1956.