On estimating survival and hazard ratios using one time-scale when cohort data have multiple time-scales: a simulation study

Nurgul Batyrbekova¹,²*, Hannah Bower³, Paul Dickman¹, Robert Szulkin², Paul C. Lambert¹,⁴ and Therese M-L. Andersson¹

*Correspondence: nurul.batyrbekova@ki.se
¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
Full list of author information is available at the end of the article

Abstract

Background: When estimating survival functions and hazard ratios during the analysis of cohort data, we often choose one time-scale, such as time-on-study, as the primary time-scale, and include a fixed covariate, such as age at entry, in the model. However, we rarely consider the possibility of simultaneous effects of multiple time-scales on the hazard function.

Methods: In a simulation study, within the framework of flexible parametric models, we investigate whether relying on one time-scale and fixed covariate as proxy for the second time-scale is sufficient in capturing the true survival functions and hazard ratios when there are actually two underlying time-scales.

Result: We demonstrate that the one-time-scale survival models appeared to approximate well the survival proportions, however, large bias was observed in the log hazard ratios if the covariate of interest had interactions with the second time-scale or with both time-scales.

Conclusion: We recommend to exercise caution and encourage fitting models with multiple time-scales if it is suspected that the cohort data have underlying non-proportional hazards on the second time-scale or both time-scales.

Keywords: multiple time-scales; flexible parametric models; simulation study; bias; survival; hazard ratios

Introduction

Survival analysis of time-to-event data can include several time-scales, such as time since diagnosis, treatment duration, attained age or calendar time. Multiple time-
scales can be seen in situations, where the event rate may vary along more than one time-scale. For example, breast cancer incidence depends on attained age as well as time since first childbirth; in many chronic diseases such as diabetes, the risk of complications is contingent on time since diagnosis and attained age; the risk of getting an infection in intensive care unit may depend on length of stay and calendar time [1, 2].

When analysing cohort data, we may be aware of the presence of two or more time-scales that may simultaneously affect the rates of an outcome. Time-on-study and attained age are the most frequently encountered time-scales. However, the joint effects of both time-scales are rarely considered. Instead, one time-scale, such as time-on-study, is often chosen as the main time-scale for the baseline hazard, while the effect of the second time-scale is not modelled, rather a time-fixed covariate, such as age at entry is included in the model.

The usual approach to jointly modelling two time-scales is to treat the main time-scale as continuous, and split the data into very short intervals along the second time-scale, thus assuming constant effects within each interval. The obtained data can then be analysed with a Cox model, or a parametric model. The flexible parametric models have become a popular alternative in analysing survival data, as they can capture complex hazard shapes, and enable smooth predictions for relative and absolute effects [3, 4, 5]. A Poisson regression model can also be used if the observations are further split along the first time-scale in addition to time-intervals on the second time-scale. This may lead to very fine splitting of the observed data to uphold the assumption of the constant hazard rate within each interval.

In large population-based register studies, splitting observations into very short intervals along two or more time-scales can become unmanageable with current technology. Time-splitting can further exacerbate the challenges of fitting very complex models. The absence of a well documented standard software to jointly model multiple time-scales can be an additional deterrent to fitting multiple time-scales models. These limitations play a role in our preference to model how the rate changes as a function of only one time-scale. Unfortunately, we rarely consider the potential bias in estimates from a misspecified model that ignores simultaneous effects of more than one time-scale.
In this study, within the framework of flexible parametric models, we aim to investigate whether modelling one time-scale while adjusting for a constant proxy for the second time-scale suffices in different situations when the hazard rates in the data depend on two time-scales. In situations where these models appear to be insufficient, we also aim to assess bias in survival and hazard ratio estimates. Assessment of choosing an optimal time-scale is out of scope of this study as there has been a considerable attention paid to the choice of a time-scale when analysing cohort data [6, 7, 8, 9], as well as combining time-scales into a unique optimal single time-scale [10, 11, 12].

Methods

Models with one or more time-scales

At least one time origin is inherent in the analysis of time-to-event data. Examples include the start of study, onset of disease or date of birth. It is common to model the log hazard function as a function of the time-scale of interest when analysing time-to-event data. The frequency of an event of interest from the time origin can dictate which time-scale(s) to use when modelling the baseline hazard function. For example, in cancer patients, the rate of deaths increases from onset of the disease, making the time from onset, or rather its proxy, the time from diagnosis, as the time-scale of interest. However, attained age, which is the time from birth, also plays an important role in the rates of death. We cannot know for certain whether one time-scale should be preferred over the other in the analysis of time-to-event data as there are many other risk factors that can influence the time at risk. Conventionally, one of the two time-scales is chosen as the primary time-scale in the baseline hazard function without further consideration that modelling both time-scales may help in capturing the true hazard function.

A general model for the log hazard function with only one time-scale, \( t_1 \), such as time from diagnosis, can be expressed as:

\[
\log(h(t_1; \gamma, \beta)) = \log(h_0(t_1; \gamma)) + \mathbf{x} \beta = g(t_1; \gamma) + \mathbf{x} \beta, \tag{1}
\]

where \( h_0 \) is the baseline hazard function, \( \gamma \) is the corresponding parameter vector and \( \mathbf{x} \) is a vector of covariates with associated log hazard ratios \( \beta \). In the Cox model,
the baseline hazard is not modelled directly due to the use of the partial likelihood\(^\text{1}\), whereas in the Poisson regression it is common to model the behaviour of\(^\text{2}\) the baseline hazard by categorising \(t_1\) through function \(g\) into time-intervals, say\(^\text{3}\) monthly or yearly intervals. If a continuous format of the time-scale is preferred then\(^\text{4}\) \(g\) can be any polynomial function or a spline function, which allows for a smooth\(^\text{5}\) non-linear function.

In situations where it is known that the hazard rate may depend on two time-scales, such as time from diagnosis and attained age, the usual approach is to include\(^\text{6}\) the time from diagnosis, \(t_1\), in the baseline hazard and use age at entry, \(a_0\), as a\(^\text{7}\) constant proxy for the second time-scale, as shown in the following model:

\[
\log(h(t_1; \gamma, \beta^*)) = \log(h_0(t_1; \gamma)) + x^*\beta^* = g(t_1; \gamma) + f(a_0; \beta_{a_0}) + x\beta. \tag{2} \text{ (3)}
\]

In this model \(a_0\) is part of the covariates \(x^*\), and the relative effect of \(a_0\), modelled\(^\text{8}\) by function \(f\), is part of the vector of the log hazard ratios \(\beta^*\). Similar to \(g\), function\(^\text{9}\) a smoothing spline function. To allow for time-dependent effects of age at diagnosis\(^\text{10}\) and of any \(x_p\) (\(p = 1, \ldots P\)) covariate in \(x\), model (2) can include interaction terms\(^\text{11}\) with \(t_1\):

\[
\log(h(t_1; \gamma, \beta_{a_0}, \gamma_{a_0}, \beta, \gamma_{x_p})) = g(t_1; \gamma) + f(a_0; \beta_{a_0}) + r(g_{int}(t_1)f_{int}(a_0); \gamma_{a_0}) + x\beta + \sum_{p=1}^{P} g_{x_p}(t_1; \gamma_{x_p})x_p. \tag{4} \text{ (5)}
\]

where \(P\) is the number of covariates, excluding age, with time-dependent effects.

Now, a model that includes both time from diagnosis, \(t_1\), and attained age, \(t_2\), in the baseline hazard function, \(h_0\), assuming proportional hazards, can be written as:

\[
\log(h(t_1, t_2; \gamma, \psi, \beta)) = \log(h_0(t_1, t_2; \gamma, \psi)) + x\beta = g(t_1; \gamma) + f(t_2; \psi) + x\beta. \tag{4} \text{ (5)}
\]

Furthermore, model (4) can also incorporate interaction between two time-scales \(t_1\) and \(t_2\):

\[
\log(h(t_1, t_2; \gamma, \psi, \tau, \beta)) = g(t_1; \gamma) + f(t_2; \psi) + r(g_{int}(t_1)f_{int}(t_2); \tau) + x\beta. \tag{5} \text{ (6)}
\]
If necessary, this model can be extended to include more time-scales, and interaction between time-scales and other covariates.

Depending on the origin of the time-scales, often one time-scale can be expressed in terms of the other. This may not always be true if the time-scale is defined in different units to time units. For example, failure rates in vehicles may be primarily measured in terms of the mileage rather than age, whereas for aircraft the time-scales can include the number of flights, the time in air and calendar time. However, in the medical setting, such as our example, using the time of diagnosis as the time origin, and \( a_0 \) as an offset term, we can specify attained age, \( t_2 \), in terms of \( t_1 \) and \( a_0 \), therefore transforming model (4) into

\[
\log(h(t_1, t_2; \gamma, \psi, \beta)) = g(t_1; \gamma) + f(t_1 + a_0; \psi) + x \beta. \tag{6}
\]

Non-proportional hazards as well as interactions between time-scales can be accommodated by including interactions between the covariates and the time-scales:

\[
\log(h(t_1, t_2; \gamma, \psi, \beta, \gamma_p, \psi_p)) = g(t_1; \gamma) + f(t_1 + a_0; \psi) + \gamma x + \sum_{p=1}^{P} \gamma_p x_p + \sum_{q=1}^{Q} \psi_q x_q. \tag{7}
\]

In this study, within the framework of flexible parametric models, our focus is on errors in predictions when fitting models as shown in equations (2), (3) with one time-scale while adjusting for a second time-scale as a constant covariate, when the true underlying rates depend on multiple time-scales as shown in equations (6), (7).

Flexible parametric models

Flexible parametric survival models (FPMs) are parametric models that use restricted cubic splines to model the baseline hazard function as well as time-dependent effects on the log hazard scale or log cumulative hazard scale [3, 4, 5]. A flexible parametric model on the log hazard scale with one time-scale that includes interactions with time, is written as:

\[
\log(h(t; \gamma_0, k_0, \beta, \gamma_p, k_p)) = s(\log(t); \gamma_0, k_0) + x \beta + \sum_{p=1}^{P} x_p s(\log(t); \gamma_p, k_p), \tag{8}
\]
where $h$ is a hazard function over time-scale $t$, $\mathbf{x}$ is a vector of covariates with associated log hazard ratios $\beta$, with additional $p^{th}$ time-dependent effect for covariate $x_p$, with $p = 1, \ldots, P$. The baseline hazard is represented by the restricted cubic spline function, $s(\log(t); \gamma_0, k_0)$, with knot location vector $k_0$, and associated coefficient vector $\gamma_0$ from the spline expansion [5]. Users need to determine the number of knots for the spline expansion of $\log(t)$ and for each of the spline expansions for the time-dependent effects. Interactions between the covariates and the restricted cubic spline function of $\log(t)$ is omitted to fit a model with proportional hazards. It should be noted that it is common to model $\log(t)$ in FPMs on the log hazard scale, but other functional forms of time $t$ can be used as well.

**Simulation strategy**

This simulation study was conducted to evaluate bias in estimates of survival functions and hazard ratios when modelling hazard rates using FPMs with one time-scale given that the true hazard rates depend on two time-scales. For this purpose, we simulated cohort data loosely based on haematological cancer patients who relapsed after transplantation, and were followed from relapse until death. We simultaneously simulated two time-scales – time since relapse and time since initial transplantation. In the interest of parsimony, we chose to exclude calendar time and attained age, which should not be ignored in the clinical study of death rates in these patients. To the simulated data we fitted FPMs with time since relapse as the time-scale, and used the time since transplantation until relapse as a constant proxy for the time since transplantation, while also allowing for possible non-proportional hazards.

**Simulation design and data generation**

We simulated 1000 datasets with a sample size of 1000 from each of the seven data-generating mechanisms (DGMs) with and without interaction terms with time-scales presented in Table 1. Each data-generating mechanism simulated time-to-event data based on two time-scales: the first time-scale, $t$, time (up to 10 years) from relapse until death or censoring, and the second time-scale, $t + c_0$, time from transplantation, where $c_0$ represents the time from transplantation until relapse, and was generated from a uniform distribution $U(0, 5)$. Notation $c_0$ for offset is used instead of $a_0$ from equations (2)-(7) to make it clear that these offsets are
from different time-scales. We used Bernoulli(0.8) for simulation of the single binary covariate \( x \). The chosen probability distributions may not be realistic within the context of relapsed patients after transplantation, but they provide a useful foundation for the purpose of this study.

Different DGMs with and without interactions with time-scales were chosen to reflect the wide spectrum of situations that we usually encounter when analysing cohort data. Scenarios \( S_1, S_2 \) generated data with proportional hazards for \( x \); \( S_3, S_4 \) simulated interaction between \( x \) and \( t \), while \( S_5, S_6 \) simulated interaction between \( x \) and \( t + c_0 \); \( S_7 \) generated data with both interaction terms: between \( x \) and \( t \), and between \( x \) and \( t + c_0 \). Of the seven data-generating mechanisms, three scenarios \( (S_2, S_4, S_6) \) also simulated interactions between the two time-scales.

For illustration of the DGMs, Figure 1 displays the survival proportions and hazard rates for \( x = 0, 1 \) as well as the log hazard ratios (\( x = 1 \) versus \( x = 0 \)) during follow-up time \( t \), which is the time from relapse until death or censoring, given relapses occurred at \( c_0 = 0.5, 4 \) years after transplantation, respectively. For scenarios \( S_5, S_6 \), it should be noted that the log hazard ratios are not technically independent of the first time-scale \( t \) due to the mathematical relationship between the first time-scale and the second time-scale (\( t_2 - t_1 = (t + c_0) - t = c_0 \)). This can be seen in declining log hazard ratios over follow-up time \( t \).

Analyses of simulated data

To each of the simulated datasets, we fitted flexible parametric models on the log hazard scale with one time-scale, \( t \), while adjusting for a constant \( c_0 \). For the restricted cubic spline expansion of \( \log(t) \), we chose six knots (five degrees of freedom), and three knots (two degrees of freedom) for the restricted cubic splines of the time-dependent effects. The choice of the number of knots was not of interest in our analyses, and according to previous studies, estimates from FPMs are not sensitive to the number of knots as long as sufficient number of knots is selected \( [14, 15, 16] \). Furthermore, the knots positions, as is most common, were chosen according to the equally-spaced centiles of the distribution of event times, with the boundary knots at the first and last event times. The effect of \( c_0 \) was also modelled with restricted cubic splines, with three degrees of freedom.
For each two-time-scale DGM we aimed to have a conceptually corresponding model with one time-scale that might be a sufficient model to capture the effects of both time-scales. Table 2 provides formulae for the fitted models and indication of correspondence with the DGMs.

Survival proportions for \( x \) and log hazard ratios were predicted from each fitted model over time-scale \( t \) for specific values of \( c_0 \). Average values of bias (the mean difference between the predicted values and true values) and 95% Monte-Carlo confidence intervals (MC CI) were evaluated at \( t = 1, 3, 5 \) years and \( c_0 = 0.5, 4 \) years.

When analysing real-world data, the underlying data generating mechanism is unknown, and the AIC [17] and the BIC [18] are often used to guide in model selection. In order to mimic this model fitting procedure we also estimated AIC and BIC for each fitted model. We calculated the percentage of times the fitted models showed lowest AIC or BIC relative to other models when fitted to the same dataset from each DGM. We used the number of events in calculation of the BIC [19].

All the results from the analyses including interactive figures with average survival predictions and average log hazard ratios for all follow-up time \( t \) for each fitted model for each simulating scenario can be found in the Supplementary Material [1].

For reference, to mirror the true models used for simulating data, we fitted flexible parametric models that jointly modelled both time-scales using three degrees of freedom for \( \log(t) \) and for \( t + c_0 \). Since these models are close approximations of the true models, they are only presented in the Supplementary Material.

### Results

Figures 2 - 4 present the bias in predicted survival functions and the log hazard ratios at follow-up times \( t = 1, 3, 5 \) (years since relapse) and \( c_0 = 0.5, 4 \) (time of relapse after transplantation) for the fitted models for the DGMs.

**Scenarios with proportional hazards**

For two simpler scenarios \( S_1, S_2 \), which simulated data with proportional hazards for \( x \), the absolute bias was low (<0.06) in the survival estimates from all fitted models.

[1] Supplementary material may also be found in the online version of this article.
models. In particular, very low absolute values (<0.005) were seen in fitted models. Larger absolute bias was observed in the log hazard ratios, where only models M1 and M2, with proportional hazards for x, showed close estimates to the true values (absolute bias <0.005). These models were aimed to correspond to S1 and S2, respectively. The AIC and BIC were both seen to also support models M1, M2 most of the time in respect to these scenarios, as shown in Table 3.

Scenarios with interactions with the first time-scale

For scenarios S3, S4, which generated data with interaction between x and the first time-scale t, the largest bias in survival proportions and the log hazard ratios was seen in the proportional hazards models. For example, model M1 in scenario S4 showed a bias of -0.08 (MCSE 0.001) in survival proportions for x = 0 at t = 1, c0 = 0.5 and 0.39 (MCSE 0.0026) for the log hazard ratio at t = 5. In contrast, model M4, which conceptually corresponds to S4, showed the lowest absolute bias in survival proportions in both S3 and S4 (<0.003, <0.006, respectively). These were closely followed by estimates from models M3 and M7. Very low bias (<0.05) was also observed in the log hazard ratios from models M3, M4 and M7. Furthermore, model M4 showed the lowest AIC 92.2% of the time compared to other models when fitting data generated by S4, and model M3 had the lowest AIC in over 67% of the fitted samples from S3.

Scenarios with interactions with the second time-scale

In comparison to other scenarios, larger absolute bias was seen in both survival proportions and the log hazard ratios from all fitted models for scenarios S5, S6 which simulated interactions with the second time-scale t + c0. Models M5, M6 that included interaction x · c0, and conceptually corresponded to scenarios S5, S6, respectively, showed low bias (<0.09) in survival predictions. However, model M7 that included both interaction x · log(t) and interaction x · c0, showed the lowest absolute bias (<0.05) in survival predictions. Furthermore, only model M7 was seen to have the closest estimates to the true log hazard ratios, particularly for t < 5. However, bias increased with increasing follow-up time, as shown in Figures 10.2, 11.2 in the Supplementary Material. In addition, model M7 showed the lowest
AIC in 83.7% of the fitted samples from $S_5$, and only in 19.4% of the fitted samples$^1$ from $S_6$.  

Scenario with interactions with both time-scales

For scenario $S_7$, which included both interaction terms between $x$ and the two time-scales, model $M7$ was conceptually the closest model. This model had the lowest absolute bias ($<0.02$) in survival predictions compared to other models. It also showed the closest log hazard ratios to the true values within the first five years of follow-up time, but bias increased with increasing follow-up time beyond five years as shown in Figure 12.2 in the Supplementary Material. The remaining fitted models had large absolute bias in both survival proportions and the log hazard ratios (>0.1, >1.1, respectively). The lowest AIC was seen for model $M7$ in 83.75% of the fitted 1000 datasets generated by scenario $S_7$.  

Discussion

The purpose of this simulation study was to assess the performance of fitted models with one time-scale (time-on-study) while incorporating time-of-entry on the second time-scale as a covariate when in reality data depend jointly on both time-scales. To this end, we simulated seven scenarios where we specified two time-scales, and evaluated bias in seven fitted flexible parametric models for each scenario. Although we focused on fitting flexible parametric models with one time-scale to the simulated data, we expect similar results from fitting Cox or Poisson regression models provided the data are split into very short intervals and a smooth function fitted to these intervals.  

The majority of the fitted models with one time-scale that approximately corresponded to the data-generating mechanisms, showed on average very close estimates to the true survival proportions and hazard ratio estimates, and most of the time were seen to have lowest AIC. Models that included an interaction between main time-scale and the constant from the second time-scale ($c_0 \cdot \log(t)$), while controlling for proportional hazards or non-proportional hazards for $x$ showed most robust results if data had proportional or non-proportional hazards for $x$ on the first time-scale, respectively.  

All models underperformed when estimating the log hazard ratios from data with interaction between $x$ and the second time-scale $t + c_0$. It is highly likely that a
researcher encountering data similar to data from $S_5, S_6$, may choose models that\(^1\) include an interaction term between $x$ and the main time-scale, such as models $M3, M4$ or $M7$, which include such an interaction. However, this can lead to misleading\(^3\) interpretation that the variation in the effect of $x$ on the hazard rates is solely due to\(^4\) the main time-scale ($t$) rather than the underlying second time-scale ($t + c_0$). Even\(^6\) though model $M7$ that includes both interaction $x \cdot \log(t)$ and interaction $x \cdot c_0$ was\(^6\) seen to have very close survival predictions to the true values, the bias in the log\(^7\) hazard ratios was still visibly large, especially for follow-up time beyond five years.\(^8\) We can never be safe from such pitfalls in the analysis of real-world data. With our\(^9\) simulation study, we hope to raise awareness of such pitfalls and direct attention\(^10\) towards more nuanced analysis with cautious approach to modeling time-scales. It\(^11\) should be noted that the simulated scenarios were designed to have joint strong\(^12\) effects from both time-scales, and therefore we do not expect to observe as severe\(^13\) bias in situations with less strong effects.

As in the majority of simulation studies, we only looked at limited number of\(^16\) scenarios, and made simplifying assumptions. In particular, we simulated data using\(^17\) only two time-scales with one type of functional form for both time-scales and\(^18\) interactions with time-scales. It would be of interest to investigate the performance\(^19\) of models fitted to the data generated using different time-scales with different\(^20\) functional forms as well as different lengths of follow-up time. In this study we\(^21\) focused on the estimation of the effect of $x$, and did not explore the estimation of\(^22\) the effect of $c_0$, which is of great interest in clinical practice. Furthermore, we did\(^23\) not investigate possible situations of smaller population size.

The growing complexity of disease progression and increasing length of follow-up\(^25\) time in chronic disease and cancer patients require more advanced survival models\(^26\) that account for outcomes on multiple time-scales and can provide more reliable\(^27\) estimates with respect to different time-points on either of the time-scales. Unfortunately, it is rarely possible to know with confidence whether the hazard rates depend\(^29\) on more than one time-scale even if the cohort data may contain other variables\(^30\) measured in units of time. If empirical evidence suggests that the hazard rates are\(^32\) dependent on multiple time-scales, then users are encouraged to use models with\(^33\) multiples time-scales and, if necessary, allow for non-proportional hazards on either
of the time-scales. Using the AIC or the BIC may be helpful in selecting the most appropriate model.

Conclusion

Through this study, we aimed to demonstrate the importance of careful consideration when modelling with one time-scale, especially when cohort data with two time-scales have interactions between covariates and at least one of the time-scales. When modelling hazard rates dependent on two time-scales, where one time-scale can be expressed as a function of the other time-scale and a constant offset, the flexible parametric models with one time-scale appear to approximate well the survival proportions and log hazard ratios, provided the models correctly accommodate the proportional or non-proportional hazards on the first time-scale as well as including interaction between the first time-scale and the constant $c_0$ (time-of-entry on the second time-scale). On the other hand, if the cohort data have interactions between the covariates and the second time-scale or both time-scales then caution should be exercised. Using models with one time-scale with non-proportional hazards for the main covariate $x$ and interaction $x \cdot c_0$ may provide good approximation of survival proportions but it may still lead to bias in the log hazard ratios. Users may find the AIC useful when making model choices in practice, but this is not guaranteed to select the most appropriate model [15].

Abbreviations

AIC: Akaike information criterion; BIC: Bayesian information criterion; DGM: data-generating mechanism; FPM: flexible parametric survival models; MC CI: Monte-Carlo confidence interval;

Acknowledgements

NB and RS are employed by SDS Life Science and their time within this project was funded by SDS Life Science, however, SDS Life Science had no involvement in the study.

Authors’ contributions

TA, HB and NB designed the study. NB performed the data analysis and drafted the manuscript. HB, TA, PCL, PD and RS supported the analysis, and provided revisions and improvements to the manuscript. All authors have read and approved the final manuscript.

Funding

This work was funded via the Swedish Cancer Society (grant number: 19 0102 Pj) and the Swedish Research Council (grant numbers: 2019-01965, 2019-00227).

Availability of data and materials

All the data were simulated for this study. The program files used for simulating data can be obtained from the corresponding author upon request.
1 Ethics approval and consent to participate
2 Not applicable.
3 Competing interests
4 The authors declare that they have no competing interests.
5 Consent for publication
6 Not applicable.
7 Author details
1 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. 2 SDS Life Science AB, Stockholm, Sweden. 3 Division of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. 4 Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK.
8 References
1. Albrektsen, G., Heuch, I., Hansen, S., Kvåle, G.: Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects. British journal of cancer 92(1), 167–175 (2005). doi:10.1038/sj.bjc.6602302
2. Wolkewitz, M., Cooper, B.S., Palomar-Martinez, M., Alvarez-Lerma, F., Olaechea-Astigarraga, P., Barnett, A.G., Schumacher, M.: Multiple time scales in modeling the incidence of infections acquired in intensive care units. BMC medical research methodology 16(1) (2016). doi:10.1186/s12874-016-0199-y
3. Royston, P., Parmar, M.K.B.: Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine 21(15), 2175–2197 (2002). doi:10.1002/sim.1203.
4. Royston, P., Lambert, P.C.: Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. StataCorp LP, (2011). https://EconPapers.repec.org/RePEc:tsj:spbook:fpsaus
5. Crowther, M.J., Lambert, P.C.: A general framework for parametric survival analysis. Statistics in Medicine 33(30), 5280–5297 (2014). doi:10.1002/sim.6300. https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.6300
6. Korn, E.L., Graubard, B.I., Midhune, D.: Time-to-Event Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-scale. American Journal of Epidemiology 145(1), 72–80 (1997). doi:10.1093/oxfordjournals.aje.a009034.
7. Thibault, A.C.M., Bénichou, J.: Choice of time-scale in cox’s model analysis of epidemiologic cohort data: a simulation study. Statistics in Medicine 23(24), 3803–3820 (2004). doi:10.1002/sim.2098.
8. Pencina, M.J., Larson, M.G., D’Agostino, R.B.: Choice of time scale and its effect on significance of predictors in longitudinal studies. Statistics in Medicine 26(6), 1343–1359 (2007). doi:10.1002/sim.2699.
9. Chalise, P., Chicken, E., McGee, D.: Performance and prediction for varying survival time scales. Communications in Statistics - Simulation and Computation 42(3), 636–649 (2013). doi:10.1080/03610918.2011.650259. https://doi.org/10.1080/03610918.2011.650259
10. Farewell, V.T., Cox, D.R.: A note on multiple time scales in life testing. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28(1), 73–75 (1979)
11. Oakes, D.: Multiple time scales in survival analysis. Lifetime data analysis 1(1), 7–18 (1995). doi:10.1007/bf00985253
12. Duchesne, T., Lawless, J.: Alternative time scales and failure time models. Lifetime data analysis 6(2), 157–179 (2000). doi:10.1023/a:1009616111968
13. Cox, D.R.: Regression models and life tables. Journal of the Royal Statistical Society B(34), 187–202 (1972)
14. Rutherford, M.J., Crowther, M.J., Lambert, P.C.: The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. Journal of Statistical Computation and Simulation
1 and Simulation 85(4), 777–793 (2015). doi:10.1080/00949655.2013.845890.

https://doi.org/10.1080/00949655.2013.845890

15. Bower, H., Crowther, M.J., Rutherford, M.J., Andersson, T.M.-L., Clements, M., Liu, X.-R., Dickman, P.W., Lambert, P.C.: Capturing simple and complex time-dependent effects using flexible parametric survival models. A simulation study. Communications in Statistics - Simulation and Computation 0(0), 1–17 (2019). doi:10.1080/03610918.2019.1634201. https://doi.org/10.1080/03610918.2019.1634201

16. Syriopoulou, E., Mozumder, S.I., Rutherford, M.J., Lambert, P.C.: Robustness of individual and marginal model-based estimates: A sensitivity analysis of flexible parametric models. Cancer Epidemiology 58, 17–24 (2019). doi:10.1016/j.canep.2018.10.017

17. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723 (1974). doi:10.1109/TAC.1974.1100705

18. Schwarz, G.: Estimating the dimension of a model. Ann. Statist. 6(2), 461–464 (1978). doi:10.1214/aos/1176344136

19. Volinsky, C.T., Raftery, A.E.: Bayesian information criterion for censored survival models. Biometrics 56(1), 256–262 (2000)

Figures

Figure 1 Survival proportions, hazard rates and log hazard ratios of the seven data generating mechanisms over time since relapse given relapses occurred at $t_0 = 0.5, 4$ years after transplantation. Note that the y-scales differ across scenarios.

Figure 2 Bias in survival proportions from fitted models for $x = 0$ for $t = 1, 3, 5$ years since relapse given relapse times at $t_0 = 0.5, 4$ years after transplantation. Note that the y-scales differ across scenarios.

Figure 3 Bias in survival proportions from fitted models for $x = 1$ for $t = 1, 3, 5$ years since relapse given relapse times at $t_0 = 0.5, 4$ years after transplantation. Note that the y-scales differ across scenarios.

Figure 4 Bias in log hazard ratios from models for $t = 1, 3, 5$ years since relapse given relapse times at $t_0 = 0.5, 4$ years after transplantation. Note that the y-scales differ across scenarios.
Table 1 Data-generating mechanisms for 1000 simulated datasets with 1000 observations with two time-scales: first time-scale, \( t \), and second time-scale, \( t + c_0 \). \textit{Int} - interaction

| DGM     | non PH-x on \( t_1 \) | non PH-x on \( t_2 \) | Int \( t_1 \cdot t_2 \) | Formula                                                                 |
|---------|------------------------|------------------------|-------------------------|-------------------------------------------------------------------------|
| \( S_1 \) | 7                      |                        |                         | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x \) |
| \( S_2 \) | 8                      | x                      |                         | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x - 0.1 \log(t) \cdot (t + c_0) \) |
| \( S_3 \) | 9                      | x                      |                         | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x - 0.2x \log(t) \) |
| \( S_4 \) | 10                     | x                      |                         | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x - 0.2x \log(t) - 0.1 \log(t) \cdot (t + c_0) \) |
| \( S_5 \) | 11                     | x                      |                         | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x - 0.2x(t + c_0) \) |
| \( S_6 \) | 12                     | x                      | x                      | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x - 0.2x(t + c_0) - 0.1 \log(t) \cdot (t + c_0) \) |
| \( S_7 \) | 13                     | x                      | x                      | \( \log(h(t; x)) = -0.5 + 0.1 \log(t) - 0.03 \log(t)^2 - 0.1(t + c_0) + 0.003(t + c_0)^2 - 0.3x - 0.2x \log(t) - 0.2x(t + c_0) \) |

Table 2 Flexible parametric models with one time-scale used to fit the simulated data with two time-scales. \textit{Int} - interaction.

| Fitted model | Corresponds to \( \text{DGM} \) | non PH-x on \( x \cdot c_0 \) | non PH-x on \( c_0 \) | Int \( x \cdot t \) | Formula                                                                 |
|--------------|---------------------------------|-------------------------------|------------------------|----------------|-------------------------------------------------------------------------|
| M1 \( S_1 \) | 26                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x \)              |
| M2 \( S_2 \) | 27                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x + \psi(s_2(c_0) \cdot \log t) \) |
| M3 \( S_3 \) | 28                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x + s_3(x \cdot \log t) \) |
| M4 \( S_4 \) | 29                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x + s_3(x \cdot \log t) + \psi(s_2(c_0) \cdot \log t) \) |
| M5 \( S_5 \) | 30                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x + \gamma(x \cdot s_2(c_0)) \) |
| M6 \( S_6 \) | 31                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x + \gamma(x \cdot s_2(c_0)) + \psi(s_2(c_0) \cdot \log t) \) |
| M7 \( S_7 \) | 32                              |                               |                         |                | \( \log(h(t; x)) = s_1(\log(t)) + s_2(c_0) + \beta_1 x + s_3(x \cdot \log t) + \gamma(\beta_3 s_2(c_0)) \) |
Table 3 Percentage of times the fitted models showed lowest AIC or BIC relative to other models when fitted to the same dataset from each data-generating mechanism. Cells in bold represent the highest percentage.

| Fitted models | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ |
|---------------|-------|-------|-------|-------|-------|-------|-------|
|               | AIC, % | BIC, % | AIC, % | BIC, % | AIC, % | BIC, % | AIC, % | BIC, % |
| M1            | 65.9   | 99.8  | 0.4   | 27.2  | 8.4   | 73.1  | 0      | 25.8  |
| M2            | 10.8   | 0     | 76    | 72.7  | 1.3   | 6.5   | 51.1   | 0      |
| M3            | 9.9    | 0.2   | 0     | 0     | 67.6  | 26.9  | 0      | 3.2   |
| M4            | 1.1    | 13    | 0     | 0     | 11.8  | 0     | 92.2   | 39.4  |
| M5            | 8.6    | 0.1   | 0.7   | 0     | 0.5   | 2.9   | 0.5    | 7.9   |
| M6            | 1.7    | 10.5  | 0.1   | 0.1   | 1.3   | 0     | 0.1    | 28    |
| M7            | 2      | 13    | 0     | 0     | 10.1  | 0     | 0      | 83.7  |

Additional Files
Additional file 1
16A HTML file containing all the results from the data analyses with additional figures and tables.