Modelling multiple time-scales with flexible parametric survival models

Nurgul Batyrbekova (✉ nurgul.batyrbekova@ki.se)

Karolinska Institutet

Hannah Bower

Karolinska Institutet

Paul W. Dickman

Karolinska Institutet

Anna Ravn Landtblom

Karolinska Institutet

Malin Hultcrantz

Memorial Sloan-Kettering Cancer Center

Robert Szulkin

SDS Life Science AB

Paul C. Lambert

University of Leicester

Therese M-L. Andersson

Karolinska Institutet

Research Article

Keywords: multiple time-scales, flexible parametric survival models, time-varying covariate, matched cohort, cohort studies, epidemiological methods

Posted Date: January 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1268869/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Modelling multiple time-scales with flexible parametric survival models

Nurgul Batyrbekova^{1,2*}, Hannah Bower³, Paul W. Dickman¹, Anna Ravn Landtblom^{4,5}, Malin Hultcrantz^{4,6}, Robert Szulkin^{1,2}, Paul C. Lambert^{1,7} and Therese M-L. Andersson¹

*Correspondence:

nurgul.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: There are situations when we need to model multiple time-scales in survival analysis. A usual approach in this setting would involve fitting Cox or Poisson models to a stacked time-split dataset. However, this leads to large datasets and can be computationally intensive, especially if interest lies in displaying how the hazard rate or survival change along multiple time-scales continuously.

Methods: We propose to use flexible parametric survival models on the log hazard scale as an alternative method when modelling data with multiple time-scales. By rewriting one of the time-scales as a function of the other time-scale, there is no need to split the data into categories of time along one or both of the time-scales.

Result: Through case-studies we demonstrate the usefulness of this method and provide examples of graphical representations of estimated hazard rates and survival proportions. The model gives nearly identical results to using a Poisson model, without requiring time-splitting.

Conclusion: Flexible parametric survival models are a powerful tool for modelling multiple time-scales. This method does not require splitting the data into small time-intervals, and therefore saves time, helps avoid technological limitations and reduces room for error.

Keywords: multiple time-scales; flexible parametric survival models; time-varying covariate; matched cohort; cohort studies; epidemiological methods

Introduction

In survival analysis, the event rates may depend on multiple time-scales simultaneously, such as time-on-study, attained age, time since disease onset, etc, all with

different time-origins, such as start date of the study, birth, onset of disease, etc. To fit survival models with multiple time-scales, it is standard practice to split the data into short time-intervals along the relevant time-scales, and fit a Cox model or a Poisson generalised linear model with categories for the time-intervals [1]. Using the Cox model may not be the best approach if we are interested in modelling the rates over the multiple time-scales, since the effect of the primary time-scale is not estimated. Furthermore, fitting either the Cox or Poisson model to the time-split data is based on the assumption of piecewise constant hazard rates within the time-intervals. Fractional polynomials [2] or splines [3] can be used instead to obtain smooth estimates of the baseline hazard function in the Poisson model. Even so, this requires splitting the data into short intervals which often leads to very large datasets, thus adding to the challenges of fitting computationally burdensome models.

In this article, we propose using flexible parametric survival models (FPMs) [4, 5] as an alternative approach when fitting hazard models with multiple time-scales. This method allows for modelling multiple time-scales as continuous functions and does not require time-splitting. FPMs with two time-scales are presented through two scenarios. In the first scenario, the second time-scale is introduced when individuals experience an intermediate event, which is considered as a time-varying covariate. In the second scenario, both time-scales are present from the start of the study but the second time-scale is relevant only for a subset of individuals in the matched cohort. We also compare the results to those from fitting the Poisson models. The choice of the optimal single time-scale or how to combine multiple time-scales into one time-scale has been widely discussed [6–11] and is not the focus of this study.

1 Notation and background

1.1 Proportional hazards models with multiple time-scales

The idea of modelling hazard rates in terms of multiple time-scales goes back to the Lexis model with two time-axes as described by Keiding (1990) [12], which was further extended to the two-way proportional hazards (PH) model by Efron (2002) [13]. In a PH model with two time-scales, t_1 and t_2 , the log hazard function can be

expressed as:

$$\log(h(t_1, t_2; \gamma_p, \gamma_s, \beta)) = p_0(f_1(t_1); \gamma_p) + s_0(f_2(t_2); \gamma_s) + \mathbf{x}\beta, \quad (1)$$

where p_0 and s_0 are the baseline hazard functions (which can be any functions) for time-scales t_1 and t_2 , respectively, with γ_p and γ_s as the corresponding parameter vectors. An intercept is included in either of the two functions. The covariates of interest are expressed as $\mathbf{x}$ with associated log hazard ratios β . The functions f_1 and f_2 are functions of the time-scale, for example $\log(t)$ or t .

Provided chosen time-scales are measured in the same units of time, and their time-origins are known then the time-scales can be expressed as functions of one reference time-scale and corresponding offset terms. For example, if t_1 is time since diagnosis and t_2 is attained age then the time-origin for t_1 is diagnosis and for t_2 it is birth. Then using age at diagnosis, a_0 , as an offset, we can write $t_2 = t_1 + a_0$, or symmetrically $t_1 = t_2 - a_0$. Thus, with t_1 as the reference time-scale, model (1) can be written as,

$$\begin{aligned} \log(h(t_1, t_2; \gamma_p, \gamma_s, \beta)) &= p_0(t_1; \gamma_p) + s_0(t_2; \gamma_s) + \mathbf{x}\beta \\ &= p_0(t_1; \gamma_p) + s_0(t_1 + a_0; \gamma_s) + \mathbf{x}\beta, \end{aligned} \quad (2)$$

And if the third time-scale of interest is the calendar time, t_3 (origin 0 CE), then we can write it as a function of t_1 , as $t_3 = t_1 + c_0$, or as a function of t_2 as $t_3 = t_2 - a_0 + c_0$, where c_0 is the calendar date of diagnosis and acts as another offset term. To demonstrate further the symmetry of model (2) with additional time-scale, we can express the model with three time-scales t_1, t_2, t_3 as a function of the reference time-scale t_2 and offset terms a_0, c_0 ,

$$\begin{aligned} \log(h(t_1, t_2, t_3; \gamma_p, \gamma_s, \gamma_q, \beta)) &= p_0(t_1; \gamma_p) + s_0(t_2; \gamma_s) + q_0(t_3; \gamma_q) + \mathbf{x}\beta \\ &= p_0(t_2 - a_0; \gamma_p) + s_0(t_2; \gamma_s) + q_0(t_2 - a_0 + c_0; \gamma_q) + \mathbf{x}\beta, \end{aligned} \quad (3)$$

1.1.1 Non-proportional hazards models with multiple time-scales

Models (1) - (3) can be extended further to include non-proportional hazards, i.e. interactions between covariates and one or more of the time-scales. For example, model (1) with interaction between covariates x_l ($l = 1, \dots, L$) and time-scale t_1 is

written as:

$$\log(h(t_1, t_2; \gamma_p, \gamma_s, \beta)) = p_0(t_1; \gamma_p) + s_0(t_2; \gamma_s) + \mathbf{x}\beta + \sum_{l=1}^L x_l p_l(t_1; \gamma_l), \quad (4)$$

Additionally, interactions between time-scales can be included, which means that the effect of the first timescale on the outcome can differ along the second timescale. In further discussions and examples, we will focus on fitting PH models with two time-scales using flexible parametric survival models and compare them to Poisson models.

1.2 Flexible parametric survival models with multiple time-scales

Flexible parametric survival models (FPMs) were introduced by Royston and Parmar (2002) [4] and have been further developed by others [5, 14, 15]. FPMs on the log-hazard scale use a smoothing function for the baseline hazard in a form of restricted cubic splines, which are piecewise cubic polynomial functions that are joined at pre-specified positions (knots). Users are required to choose a number of knots for the splines, as well as the position of the knots. Sensitivity analyses have shown that estimates produced by FPMs are robust to the number and placement of knots [16].

As described in Section 1.1, if the time-scales of interest can be expressed in terms of one time-scale and the offset terms then, in its simplest form, the PH FPM on the log hazard scale with two time-scales, t_1 and t_2 is written as

$$\log(h(t_1, t_2; \gamma_p, \mathbf{k}_p, \gamma_s, \mathbf{k}_s, \beta)) = p_0(t_1; \gamma_p, \mathbf{k}_p) + s_0(t_2; \gamma_s, \mathbf{k}_s) + \mathbf{x}\beta, \quad (5)$$

where the baseline hazard functions on the t_1 and t_2 time-scales are represented by the restricted cubic spline functions, $p_0(t; \gamma_p, \mathbf{k}_p)$ and $s_0(t; \gamma_s, \mathbf{k}_s)$, respectively. The knot location vectors, $\mathbf{k}_p$ and $\mathbf{k}_s$ are either chosen by default at equally spaced centiles over the distribution of the event-times or determined by the user. As in the general approach presented previously, other functional forms of t_1 , t_2 , such as $\log(t_1)$ and $\log(t_2)$ can also be used within the FPM. Also this model can be easily extended to include interactions between the time-scales as well as non-proportional hazards for the covariates of interest.

1.3 Poisson model with multiple time-scales

Models of type (1) - (3) can be adapted within the Poisson framework [13, 17, 18]. The user is required to split the data along the relevant time-scales into short intervals to better support the assumption of constant hazard rates within the intervals, which can come at the cost of high computational burden. The log of person-time within each interval is included in the model as an offset. For consistency with the FPM approach, in this study we use restricted cubic splines in Poisson models to estimate the smooth effect of the time-scales on the baseline hazard. Interactions and non-PH hazards can also be modelled.

2 Application scenarios

To illustrate models with multiple time-scales we analysed two different case-studies, using both FPMs and Poisson models. The first case-study includes a time-varying variable that introduces a second time-scale upon changing its value, and the second is a matched cohort study where the second time-scale is relevant only for the exposed subjects. In both case-studies our objective was to model mortality rates with two time-scales, and graphically represent the estimated rates and the survival proportions over different time-points on both time-scales.

2.1 Time-varying covariate

Similar to the illness-death example discussed by Iacobelli and Carstensen (2013) [18], for modelling multiple time-scales, we analysed cohort data of 2,982 individuals diagnosed during 1978-1993 with primary breast cancer in Rotterdam, who had undergone primary surgery [19, 20]. Patients were followed from the initial state *surgery* until the event of interest *death* or censoring at 10 years post surgery. Throughout the follow-up it is also known whether and when they experienced the intermediate state *relapse or metastasis* (RM). RM can therefore be treated as a time-varying covariate, and time since RM as a secondary time-scale.

In total 1,139 patients died during the total follow-up of 19,937 person-years (overall mortality rate 57 per 1000 person-years), and 1,004 of these deaths occurred after RM. During follow-up, 1,477 patients experienced RM, and the median time to RM among those with RM was 2.43 years (min = 0.1, max = 9.99).

Using the FPM and the Poisson methods, we fitted the following hazard model with two time-scales, t_1 as time since surgery to death (reference time-scale) and

$t_1 - r$ as time since RM to death with r as the time of RM,

$$\begin{aligned} \log(h(t_1, t_1 - r; \gamma_p, \mathbf{k}_p, \gamma_s, \mathbf{k}_s, \gamma_q, \mathbf{k}_q, \beta)) = & p_0(\log(t_1); \gamma_p, \mathbf{k}_p) + s_0(t_1 - r; \gamma_s, \mathbf{k}_s) \cdot I_{RM} \\ & + \beta_{RM} \cdot I_{RM} + q(age; \gamma_q, \mathbf{k}_q) + \beta_{horm} \cdot I_{horm}, \end{aligned} \quad (6)$$

where I_{RM} was used as a time-varying indicator for patients with RM, so that the effect of time since RM on the baseline hazard can only be estimated after patients have transitioned to RM. In both methods, for the restricted cubic spline functions, p_0, s_0 , we chose four knots equally placed over the distribution of event-times, and we used the same knots in both methods. An intercept of the model is included in function p_0 as part of the parameters for the reference time-scale t_1 . Furthermore, we included in the model a restricted cubic spline function of *age* (age at surgery) with six knots and an indicator I_{horm} for hormonal therapy, assuming proportional hazards for these covariates over both time-scales by not including interactions with either of the time-scales.

To fit this model, it was necessary to create multiple rows per individuals for those that experienced RM, where indicator I_{RM} changed from 0 to 1 at the time of RM. Furthermore, for the Poisson approach, we split the reference time-scale, time since surgery, into two-day-intervals, to uphold the assumption of the piecewise constant hazards. For this cohort it may be unnecessary to have such short intervals, especially in later part of the follow-up, but for other diseases it may be essential.

For the purpose of this study, we chose to keep the model simple in terms of the number of covariates, and we did not focus on the question of model-specification. Therefore, the results may not be clinically representative.

The implementation of analyses for the FPM method for this cohort using Stata code is provided in the supplementary material.

2.1.1 Graphical representations of results for model with a time-varying covariate

Both the FPM and Poisson approaches yielded almost identical coefficients and 95% confidence intervals (CI) displayed in Table 1. The spline coefficients are not interpretable on their own but they are the “building blocks” for predicting the complex shape of the hazard surface for different covariate values. For example, Figure 1 displays two panels for viewing the estimated mortality rates from the

FPM. The left panel shows the estimated mortality rates per 1000 person-years along time since surgery (time-scale t_1) for patients who did not have RM and for patients who experienced RM at 0.5, 1, 2, 3, 4, 5 years post surgery, respectively. The jumps from “No RM” at these time-points post-surgery into the RM state are represented by the straight lines. The right panel of Figure 1 provides an alternative view, along the second time-scale, time since RM (time-scale $t_2 = t_1 - r$, with r as offset as time of RM post surgery). From both panels, as from Supplementary Figure 1, we see that the mortality rates decrease with both time since surgery and with time since RM. Patients treated without the hormonal therapy also have lower mortality rates than patients who received hormonal therapy. As in Figure 1, in subsequent figures, we displayed results from the FPM only, as there was a complete overlap between the two methods of estimated mortality rates and survival proportions. Furthermore, for demonstration purposes, all the predicted values were calculated for patients who had primary surgery at age 50.

Supplementary Figure 1 shows the estimated mortality rates per 1000 person-years over both time-scales in three-dimensional view for patients who had primary surgery at age 50, received hormonal therapy and experienced RM. It can be seen that the peak of the surface with grey and orange belts (mortality rates higher than 600) is concentrated over the quadrant of one to three years since surgery and one to three years since RM. The mortality rates surface becomes more “shallow” with progression of time on both time-scales similar to what is observed in clinical practice [21].

We can also assess the time-varying effect of experiencing RM on mortality rates compared to not having RM given the same time since surgery, same age at surgery and same treatment with or without hormonal therapy. This is represented by mortality rate ratio with 95% CI over time since RM in Figure 2. From Figure 2 we observe that the relative effect of having RM increases rapidly with time since RM: from rate ratio 24.27 (95% CI: 18.75, 31.4) at time 0 since RM to rate ratio of 52.51 (95% CI: 41.2, 66.93) at 3.65 years since RM, followed by a rapid decline reaching 19.91 (95% CI: 8.02, 49.41) at 10 years since RM. To clarify further this comparison, the mortality rate for individuals at 10 years since RM is 19.91 times the mortality rate for individuals who never experience RM given these individuals have the same time since surgery, the same age and therapy. The model assumes

that the mortality rate ratio at a certain time since RM (i.e. at certain values of $t_2=t_1-r$) is the same irrespective of time since surgery (i.e. t_1) and time of RM (i.e. r) as long as the comparison is made between individuals having the same time since surgery. This means that we get the same mortality rate ratio for the effect of RM in comparison to no RM when, for example, $t_1 = 2$, $r = 1$, and $t_1 = 3$, $r = 2$. This comes from the fact that the model does not include an interaction term between RM and t_1 , nor does it include an interaction between t_1 and t_2 time-scales. For more complex models that include interactions between time-scales as well as time-dependent effects of the covariates, this is not the case, and graphical representation of the mortality rate ratio over different values of both the time-scales plays even more crucial role for understanding the disease at hand.

In addition to the mortality rates, we can also obtain survival proportions for different sub-groups along both time-scales. The 95% CI (not shown) were computed using the bias-corrected percentile bootstrap method with 1000 samples. Figure 3 provides four examples of predicted survival probabilities on either of the time-scales. For example, panel (A) shows smooth predicted survival curves over time since surgery for patients without RM and for patients with RM at 1, 2, 3, 4, 5 years post surgery. Panel (B), on the other hand, depicts the survival proportions over time since RM. Both panels supplement each other and aid in comparison of the sub-groups. In addition to the conventional representations of predicted survival in panels (A) and (B), it can be of interest to assess the survival proportions chosen for specific time-points since RM and time-points since surgery. Panel (C) displays the probability of surviving 1, 3 and 5 years after RM across different time points since surgery. It can be seen that the survival proportions are higher for patients with longer time since surgery given the same time since RM. For example, the probability to survive three years after RM for patients on hormonal therapy 0.206 (95% CI: 0.142, 0.276) and 0.465 (95% CI: 0.394, 0.544) at four and eight years since surgery, respectively. Here we continue to observe the large gap between the survival for patients treated with and without the hormonal therapy.

2.2 Matched cohort

In a matched cohort, when individuals with a certain disease or characteristic are matched to population controls, or comparators without the characteristics, using

matching variables such as age, sex and calendar year, then attained age is often the most relevant time-scale. However, time since diagnosis can also be of importance for the exposed individuals, but is not relevant for the matched comparators in the cohort. We explore this type of scenario by analysing matched cohort data of Swedish patients with myeloproliferative neoplasms (MPN). Our objective was to model mortality rates and estimate survival proportions for the MPN cases compared to population comparators over the time-scales attained age and time since diagnosis.

A detailed description of the disease and patient characteristics of the MPN matched cohort are provided elsewhere [22]. Briefly, each MPN patient at age 18 years or older was matched to four individuals from the Swedish general population based on sex, age and calendar year of diagnosis during 1987 - 2009. Furthermore, the first 30 days of follow-up were excluded to avoid surveillance bias. In total the cohort consisted of 9,164 MPN cases and 35,763 matched controls with follow-up until death or censoring on December 31, 2010 or at maximum of 10 years of follow-up. In total 5,108 deaths were observed among MPN subjects, and 11,677 among non-MPN subjects (overall mortality rates were 91 and 39 per 1000 person-years, respectively).

Similar to the previous case-study, using the FPM and Poisson methods we fitted the following hazard model,

$$\begin{aligned} \log(h(t_1, t_1 - a_0; \gamma_g, \mathbf{k}_g, \gamma_v, \mathbf{k}_v, \beta)) = & g_0(t_1; \gamma_g, \mathbf{k}_g) + v_0(t_1 - a_0; \gamma_v, \mathbf{k}_v) \cdot I_{MPN} \\ & + \beta_{MPN} \cdot I_{MPN} + \beta_{sex} \cdot I_{sex}, \end{aligned} \quad (7)$$

where t_1 is attained age (reference time-scale), $t_1 - a_0$ is time since diagnosis (second time-scale) with a_0 as age at MPN diagnosis, I_{MPN} is the indicator for the MPN cases, and I_{sex} is the indicator for sex (0 = Men, 1 = Women). Even though this model might not be the most relevant, in terms of covariates included, it is useful for demonstration purposes. In both FPM and Poisson methods, we chose five knots for both spline functions g_0, v_0 , equally placed over the distribution of event-times, where the first knot and the last knot are placed at the first and last event times,

respectively. Furthermore, for the Poisson method, we split the person-time by two-day intervals.

2.2.1 Graphical representations of results for matched cohort

As in the previous case-study, the coefficients and 95% CI displayed in Table 2 from both the FPM and Poisson approaches are almost identical. The predicted mortality rates and survival proportions were the same from both methods, hence the figures that follow represent results for attained ages 60-95 from the FPM method only.

The mortality rates surface per 1000 person-years for male MPN patients in Supplementary Figure 2 shows a characteristic steep increase in the mortality rates with progression of time on both time-scales, attained age and time since diagnosis. It can also be seen that for the same attained age, the mortality is much higher right after diagnosis than after 10 years of having the disease. For example, among men at age 80 right after the diagnosis, the mortality rate is 260 per 1000 person-years, whereas for 80 year-olds after 10 years of having the disease it is 149 per 1000 person-years. However, this mortality rate is still higher than the mortality rate for 80 year-olds after five years of having the disease (133 per 1000 person-years). These differences and similar results for other ages can be seen in both panels of Figure 4. Furthermore, the rates among women are shown to be consistently lower in both MPN and non-MPN subjects.

After initial steep decline in the first two years since diagnosis, given the same attained age and sex, the mortality rate ratio is shown to be averaging at 2.51 (95% CI: 2.33, 2.7) for MPN cases relative to non-MPN matched controls over time since diagnosis (Figure 5).

Different graphical views of survival proportions are shown in Figure 6. Panels (A) and (B) provide two alternative displays of the same survival proportions for patients with ages at diagnosis 70, 75, 80, 85 over time-scales time since diagnosis and attained age, respectively. As expected younger patients have better survival as well as women in all age groups. In panel (C), we have a cross-sectional view of panel (B), where 1-, 3-, 5-, 10-year survival proportions are plotted with respect to age at diagnosis. The dramatic decrease in survival for older age groups given the same lengths of time since diagnosis is more apparent in this figure. For example,

the 5-year overall survival is 81.07% (95% CI: 80.09, 81.99) and 58.49% (95% CI: 57.09, 59.8) for a male patient diagnosed at age 69 and 79, respectively.

3 Discussion

There are situations when it is necessary to include multiple time-scales in a hazard model. For example, both woman's age and time since first childbirth have a simultaneous impact on breast cancer incidence [23] and mortality among individuals with dementia is dependent on time since onset and attained age [24] and for chronic diseases, such as diabetes mellitus, rates of complications are dependent on person's age and duration of the disease [25]. The study design may dictate which time-scales to include in the hazard model. For example, in the illness-death framework by allowing non-Markov assumption, the baseline hazard can take into account both time from the initial state as well as time from the intermediate state. And, in the matched cohort design, it may be required that the hazard model includes additional time-scales specific to the subgroups of the cohort. A recent study has also demonstrated how multiple time-scales can be incorporated within the relative survival framework in a multistate setting [26].

An established method, such as the Poisson GLM, is a powerful tool when it comes to modeling the hazard rates with multiple time-scales. However, with longer follow-up times, and larger volumes of data, it can be computationally challenging to model with a large stacked dataset containing finely split person-time. In addition to a potentially long processing time, there is also a higher risk of making errors when it comes to splitting the person-time along multiple time-scales.

In this study we aimed to introduce and demonstrate the flexible parametric survival models that can capture complex shapes of the baseline hazard function with multiple time-scales given other time-scale(s) are expressed in terms of the reference time-scale and offsets or times of origin. We illustrated through two case-studies that both the Poisson model and the FPM provide nearly identical results, with the FPM showing greater flexibility when it comes to fitting the model and making predictions. Furthermore, we demonstrated different ways of presenting the results graphically, that can help users understand the disease better. Plotting the hazard rates and survival probabilities with respect to each time-scale separately for different time-points from the other time-scale can also be helpful to answer different

research questions in regards to the disease of interest. The implementation of all the analyses in Stata for the model with time-varying covariate is shown in the supplementary material.

There are limitations to the methods proposed in this study. The survival prediction requires numerical integration, which has not been implemented in the software used, and the confidence intervals were obtained using the bias-corrected percentile bootstrap. Future work should implement easier estimation of survival proportions as well as other methods for estimating confidence intervals with the use of the delta method [27]. Another limitation that is inherent to the FPM, is choosing knots for the spline functions. However, according to Syriopoulou et al. (2019)[28], as long as there are not too few knots, the results are robust to different choices of knots. We also did not address the question of model fitting in this study. This is the most challenging part of fitting models with multiple time-scales as there are so many aspects to consider in addition to choosing covariates, such as determining which time-dependent effects on which time-scales to include, or whether there are interactions between the time-scales. With each time-scale comes an additional dimension of all the challenges that users face with one-time-scale models.

In conclusion, by using the FPM, users can avoid splitting the survival data into fine intervals, thus reducing room for error while obtaining all the necessary estimates for inference and the graphical representation of the estimated hazard rates and probabilities. The examples shown were used to highlight the importance of fitting models with multiple time-scales, and how results from these complex models can be graphically presented.

Abbreviations

FPM: flexible parametric survival model; PH: proportional hazards; RM: relapse or metastasis; CI: confidence intervals; MPN: myeloproliferative neoplasm; GLM: generalised linear model

Authors' contributions

TA and NB designed the study. NB performed the data analysis and drafted the manuscript. HB, TA, PCL, PD, RS, ARL and MH 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) and the Swedish Research Council (grant numbers: 2019-01965, 2019-00227).

Availability of data and materials

Analysis data as well as Stata code for the model with a time-varying covariate are provided in the supplementary material.

Ethics approval and consent to participate

The study was approved by the Ethical Review Board (Dnr 2017/73-31; Dnr 2020-05339; Dnr 2005/206-31/3 with amendments 2011/809-32, 2013/1353-32 and 2014/1610-32). Informed consent was not required as we had no contact with the study individuals nor our analyses contained any personal identifiers. The waived consent was obtained from the Ethical Review Board for this study. All methods were performed with accordance with relevant guidelines and regulations.

Conflict of interest

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. TA, HB, PCL, PD, ARL, MH declare that they have no conflict of interest.

Consent for publication

Not applicable.

Author details

¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ²SDS Life Science AB, Stockholm, Sweden. ³Clinical Epidemiology Division, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden. ⁴Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden. ⁵Department of Medicine, Division of Hematology, Stockholm South Hospital, Stockholm, Sweden. ⁶Department of Medicine, Myeloma Service, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. ⁷Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK.

References

1. Efron, B.: Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve **83**(402), 414–425. doi:10.2307/2288857. 2288857
2. Royston, P., Altman, D.G.: Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **43**(3), 429–467 (1994)
3. Herndon II, J.E., Harrell Jr., F.E.: The restricted cubic spline hazard model. *Communications in Statistics - Theory and Methods* **19**(2), 639–663 (1990). doi:10.1080/03610929008830224
4. 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 **21**(15), 2175–2197. doi:10.1002/sim.1203. Accessed 2021-04-27
5. Lambert, P.C., Royston, P.: Further Development of Flexible Parametric Models for Survival Analysis **9**(2), 265–290. doi:10.1177/1536867X0900900206. Accessed 2021-04-27
6. 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). doi:10.2307/2346815
7. Oakes, D.: Multiple time scales in survival analysis. *Lifetime Data Analysis* **1**(1), 7–18 (1995). doi:10.1007/BF00985253
8. Korn, E.L., Graubard, B.I., Midthune, 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
9. Thiébaud, 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
10. 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
11. 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
12. Keiding, N.: Statistical inference in the Lexis diagram. doi:10.1098/rsta.1990.0128
13. Efron, B.: The two-way proportional hazards model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4), 899–909 (2002). doi:10.1111/1467-9868.00368

14. Crowther, M.J., Lambert, P.C.: A general framework for parametric survival analysis **33**(30), 5280–5297. doi:10.1002/sim.6300. Accessed 2021-04-27
15. Bower, H., Crowther, M.J., Lambert, P.C.: Strcs: A Command for Fitting Flexible Parametric Survival Models on the Log-hazard Scale. *The Stata Journal: Promoting communications on statistics and Stata* **16**(4), 989–1012 (2016). doi:10.1177/1536867X1601600410. Accessed 2021-07-16
16. 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. Publisher: Taylor & Francis .eprint: <https://doi.org/10.1080/03610918.2019.1634201>. Accessed 2021-04-27
17. Carstensen, B.: Age-period-cohort models for the Lexis diagram **26**(15), 3018–3045. doi:10.1002/sim.2764. 17177166
18. Iacobelli, S., Carstensen, B.: Multiple time scales in multi-state models. *Statistics in Medicine* **32**(30), 5315–5327 (2013). doi:10.1002/sim.5976
19. Foekens, J.A., Peters, H.A., Look, M.P., Portengen, H., Schmitt, M., Kramer, M.D., Brünner, N., Jänicke, F., Meijer-van Gelder, M.E., Henzen-Logmans, S.C., van Putten, W.L., Klijn, J.G.: The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients **60**(3), 636–643. 10676647
20. Sauerbrei, W., Royston, P., Look, M.: A New Proposal for Multivariable Modelling of Time-Varying Effects in Survival Data Based on Fractional Polynomial Time-Transformation **49**(3), 453–473. doi:10.1002/bimj.200610328. Accessed 2021-06-02
21. Dent, R., Valentini, A., Hanna, W., Rawlinson, E., Rakovitch, E., Sun, P., Narod, S.A.: Factors Associated with Breast Cancer Mortality after Local Recurrence. *Current Oncology* **21**(3), 418–425 (2014). doi:10.3747/co.21.1563
22. Hultcrantz, M., Björkholm, M., Dickman, P.W., Landgren, O., Derolf, Å.R., Kristinsson, S.Y., Andersson, T.M.L.: Risk for Arterial and Venous Thrombosis in Patients With Myeloproliferative Neoplasms **168**(5), 317–325. doi:10.7326/M17-0028. Accessed 2021-06-04
23. 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 **92**(1), 167–175. doi:10.1038/sj.bjc.6602302. Accessed 2021-04-27
24. Commenges, D., Joly, P., Letenneur, L., Dartigues, J.: Incidence and mortality of Alzheimer's disease or dementia using an illness-death model **23**(2), 199–210. doi:10.1002/sim.1709. Accessed 2021-04-27
25. Huo, L., Magliano, D.J., Rancière, F., Harding, J.L., Nanayakkara, N., Shaw, J.E., Carstensen, B.: Impact of age at diagnosis and duration of type 2 diabetes on mortality in Australia 1997-2011 **61**(5), 1055–1063. doi:10.1007/s00125-018-4544-z. Accessed 2021-04-27
26. Weibull, C.E., Lambert, P.C., Eloranta, S., Andersson, T.M.L., Dickman, P.W., Crowther, M.J.: A multistate model incorporating estimation of excess hazards and multiple time scales. *Statistics in Medicine* **40**(9), 2139–2154 (2021). doi:10.1002/sim.8894
27. Appendix 1: The Delta Method. In: *Applied Survival Analysis*, pp. 355–358. John Wiley & Sons, Ltd (2008). doi:10.1002/9780470258019.app1. <http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470258019.app1>
28. 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 **58**, 17–24. doi:10.1016/j.canep.2018.10.017. Accessed 2021-05-28

Figures

Figure 1: Based on Rotterdam Breast Cancer data, estimated mortality rates per 1000 person-years for breast cancer patients who were aged 50 at primary surgery and were treated with hormonal therapy (solid lines) and without hormonal therapy (dotted lines). **Left panel** shows the mortality rates over time since surgery (time-scale t_1) for patients having relapse or metastasis (RM) at different time-points since surgery as well as for non-RM patients. **Right panel** shows the mortality rates for RM patients along the time since RM (time-scale t_2).

Figure 2: Estimated mortality rate ratio and 95% confidence interval for breast cancer patients in the Rotterdam Breast Cancer data, experiencing relapse or metastasis (RM) after primary surgery versus patients without RM, over time since RM.

Figure 3: Predicted survival proportions for breast cancer patients in the Rotterdam Breast Cancer data, who were aged 50 at primary surgery and were treated with hormonal therapy (solid lines) and without hormonal therapy (dotted lines). **Panel (A)** shows the survival proportions over time since surgery (time-scale t_1) for patients having relapse or metastasis (RM) at different time-points since surgery as well as for non-RM patients. **Panel (B)** shows the survival proportions since RM (time-scale t_2) for patients having RM at different time-points. **Panel (C)** shows 1-, 3- and 5-year survival proportions since RM across time since surgery.

Figure 6: Predicted survival proportions for Swedish MPN patients where solid lines are for men and dotted lines are for women. **Panel (A)** shows the survival proportions over time since diagnosis (time-scale t_2). **Panel (B)** shows the survival proportions over attained age (time-scale t_1) for different ages at diagnosis. **Panel (C)** shows 1-, 3-, 5- and 10-year survival proportions across age at diagnosis.

Tables

Table 1: Coefficients and 95% confidence intervals from the FPM and Poisson models of the mortality rates with two time-scales (t1 as time since surgery, and t2 as time since relapse or metastasis (RM)) and a time-varying indicator for the second time-scale (Indicator for RM) for Rotterdam Breast Cancer patients. Note: the spline coefficients are not interpretable on their own but they are used for making smooth predictions of the mortality rates.

	FPM	Poisson
Indicator for RM	3.1891 (2.9314, 3.4468)	3.1895 (2.9324, 3.4466)
Age spline 1	-0.0578 (-0.1092, -0.0063)	-0.0578 (-0.1093, -0.0063)
Age spline 2	-0.0005 (-0.0009, 0)	-0.0005 (-0.0009, 0)
Age spline 3	0.0006 (-0.0002, 0.0014)	0.0006 (-0.0002, 0.0014)
Age spline 4	-0.0001 (-0.0008, 0.0005)	-0.0001 (-0.0008, 0.0005)
Age spline 5	-0.0001 (-0.0004, 0.0002)	-0.0001 (-0.0004, 0.0002)
Indicator for hormonal therapy	0.3224 (0.1439, 0.5009)	0.3224 (0.1439, 0.5009)
log(t1) spline 1	0.8853 (0.4266, 1.344)	0.8777 (0.4209, 1.3346)
log(t1) spline 2	0.7396 (0.3272, 1.1519)	0.7374 (0.3253, 1.1494)
log(t1) spline 3	-1.0513 (-1.7461, -0.3565)	-1.0489 (-1.7434, -0.3543)
t2 spline 1	0.4519 (0.1646, 0.7392)	0.4533 (0.1669, 0.7398)
t2 spline 2	0.0562 (-0.0555, 0.1679)	0.057 (-0.0545, 0.1685)
t2 spline 3	-0.0153 (-0.0711, 0.0404)	-0.0157 (-0.0714, 0.0399)
Intercept	-4.0529 (-4.3845, -3.7213)	-4.0546 (-4.386, -3.7231)

Table 2: Coefficients and 95% confidence intervals from the FPM and Poisson models of the mortality rates with two time-scales (t1 as attained age, and t2 as time since MPN diagnosis) for cohort of Swedish MPN patients diagnosed during 1987-2009 and matched comparators. Note: the spline coefficients are not interpretable on their own but they are used for making smooth predictions of the mortality rates.

	FPM	Poisson
Indicator for MPN	1.5563 (1.4543, 1.6583)	1.5585 (1.457, 1.66)
Indicator for sex (0 = Men, 1 = Women)	-0.3273 (-0.3623, -0.2923)	-0.3272 (-0.3622, -0.2922)
t1 spline 1	0.0657 (0.0508, 0.0805)	0.0657 (0.0508, 0.0805)
t1 spline 2	0.0001 (-0.0001, 0.0002)	0.0001 (-0.0001, 0.0002)
t1 spline 3	-0.0006 (-0.001, -0.0003)	-0.0006 (-0.001, -0.0003)
t1 spline 4	0.0007 (0.0004, 0.001)	0.0007 (0.0004, 0.001)
t2 spline 1	-0.4661 (-0.6089, -0.3234)	-0.4704 (-0.6126, -0.3282)
t2 spline 2	-0.0821 (-0.1315, -0.0327)	-0.0833 (-0.1326, -0.034)
t2 spline 3	0.0344 (0.0045, 0.0644)	0.0351 (0.0051, 0.065)
t2 spline 4	-0.0028 (-0.0113, 0.0058)	-0.0028 (-0.0114, 0.0057)
Intercept	-8.8858 (-9.6864, -8.0853)	-8.8855 (-9.6858, -8.0852)

Additional Files

Stata .do files containing Stata code for the analysis of the model with time-varying covariates. Supplementary figures.

Supplementary Figure 1: Estimated mortality rates surface per 1000 person-years over two time-scales for breast cancer patients in the Rotterdam Breast Cancer data, who were aged 50 at primary surgery, received hormonal therapy and experienced relapse or metastasis (RM).

Supplementary Figure 2: Estimated mortality rates surface per 1000 person-years over two time-scales for male MPN patients in Sweden.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.zip](#)
- [FigureS1.pdf](#)
- [FigureS2.pdf](#)