mexhaz: An **R** Package for Fitting Flexible Hazard-Based Regression Models for Overall and Excess Mortality with a Random Effect

Hadrien Charvat iD
National Cancer Center Japan

Aurélien Belot iD
London School of Hygiene &
Tropical Medicine

Abstract

We present **mexhaz**, an R package for fitting flexible hazard-based regression models with the possibility to add time-dependent effects of covariates and to account for a two-level hierarchical structure in the data through the inclusion of a normally distributed random intercept (i.e., a log-normally distributed shared frailty). Moreover, **mexhaz**-based models can be fitted within the excess hazard setting by allowing the specification of an expected hazard in the model. These models are of common use in the context of the analysis of population-based cancer registry data.

Follow-up time can be entered in the right-censored or counting process input style, the latter allowing models with delayed entries. The logarithm of the baseline hazard can be flexibly modeled with B-splines or restricted cubic splines of time. Parameters estimation is based on likelihood maximization: in deriving the contribution of each observation to the cluster-specific conditional likelihood, Gauss-Legendre quadrature is used to calculate the cumulative hazard; the cluster-specific marginal likelihoods are then obtained by integrating over the random effects distribution, using adaptive Gauss-Hermite quadrature. Functions to compute and plot the predicted (excess) hazard and (net) survival (possibly with cluster-specific predictions in the case of random effect models) are provided. We illustrate the use of the different options of the **mexhaz** package and compare the results obtained with those of other available R packages.

Keywords: adaptive Gauss-Hermite quadrature, excess hazard, flexible models, frailty models, time-dependent effects, C.

1. Introduction

In the context of the analysis of time-to-event data, parametric and semi-parametric hazard regression models are widely used when the interest lies in estimating the impact of covariates

on the time to occurrence of the event of interest. The semi-parametric Cox proportional hazard model is still widely used, even if its creator himself argued that parametric models should be used more often in practice, due to powerful features for in- and out-sample predictions, as well as the advantage of allowing statistical inference using maximum likelihood theory (see Reid 1994). However, using parametric regression models requires the assumption of a particular distribution for the observed survival times, and this assumption may sometimes be considered as too restrictive (e.g., constant or monotonic hazard for the exponential and Weibull distribution, respectively). One possibility to take advantage of parametric models without making unrealistic assumptions on the shape of the hazard (and on the corresponding survival) is to use flexible functions, such as fractional polynomials or regression splines. The correct modeling of the data might also require the inclusion of time-dependent effects of some of the covariates. Indeed, it has been shown in many studies that the effects of covariates such as age may vary with time since diagnosis, especially in cancer epidemiology, so that the proportional hazard assumption does no longer hold (Quantin *et al.* 1999; Bossard *et al.* 2007).

Another aspect one might have to deal with is the presence of a hierarchical structure in the data: individuals from the same cluster share common characteristics (e.g., cancer patients from the same geographical area may have similar access to therapeutical resources) so that the assumption of independence of the survival times no longer holds and taking account of this hierarchical structure is necessary for correct statistical inference. In such a case, shared frailty models (also called multilevel or mixed-effect survival models) have been shown to provide a satisfactory and convenient theoretical framework by allowing the introduction of a random effect defined at the cluster level that accounts for the inter-cluster heterogeneity (Duchateau and Janssen 2008; Wienke 2010).

In population-based cancer research, it is generally of interest to disentangle the cancer-specific mortality from the mortality from other causes, mainly because cancer patients are usually old and, as a consequence, more prone to die from diseases other than their cancer (these other diseases thus act as competing causes of death). The general principle is to assume that the observed mortality hazard can be split into two components, one representing the mortality from cancer and the other one representing the impact of other causes of death. When information on the cause of death is available for each individual, this can be achieved by estimating cause-specific mortality hazards (Putter, Fiocco, and Geskus 2007; Belot, Abrahamowicz, Remontet, and Giorgi 2010; Haller, Schmidt, and Ulm 2013). However, when using population-based cancer registry data, the cause of death is usually unavailable or inaccurate (it might even be difficult to define a cause of death for elder patients with multiple diseases). Thus specific methods have been developed (Estève, Benhamou, Croasdale, and Raymond 1990; Giorgi *et al.* 2003; Nelson, Lambert, Squire, and Jones 2007; Remontet, Bossard, Belot, and Estève 2007; Pohar Perme, Stare, and Estève 2012) that rely on the same general principle than in the cause-specific setting (i.e., the overall mortality rate is seen as the sum of two components), but it requires the additional assumption that the mortality hazard for other causes of death can be approximated by the mortality hazard of the general population (for a given set of demographic characteristics observed on each cancer patient). These methods allow the estimation of the so-called excess mortality hazard, which can be interpreted as the cancer-specific mortality hazard. The net survival, i.e., the survival that would be observed if cancer patients could only die from their cancer, can then be obtained from the estimated excess mortality hazard (Belot *et al.* 2019).

Within the R software environment (R Core Team 2021), different packages have been developed for fitting flexible hazard models based on a full or penalized likelihood framework. In the full likelihood framework, the contributed R package **flexsurv** (Jackson 2016) uses specific distributions for the survival time, including the generalized gamma and F distribution families and also spline-based models. Facilities are also proposed to fit excess hazard regression models. Another R package named **flexrsurv** (Clerc-Urmès and Grzebyk 2020) has been developed for fitting two types of flexible hazard regression models (Remontet *et al.* 2007; Mahboubi, Abrahamowicz, Giorgi, Binquet, Bonithon-Kopp, and Quantin 2011) in the excess hazard setting. It is also worth mentioning the **relsurv** package (Pohar Perme and Pavlič 2018; Pohar Perme and Stare 2006, 2007) which, although primarily aimed at non-parametric net survival estimation, can also be used to fit excess hazard regression models with either a baseline hazard described by piecewise constant functions (full likelihood framework) or with a baseline hazard left unspecified (in the same spirit as the Cox model) using an expectation-maximization algorithm for parameter estimation (Pohar Perme, Henderson, and Stare 2009). However, neither **flexsurv**, **flexrsurv** nor **relsurv** has the possibility to account for correlated survival times. The package **rstpm2** (Clements, Liu, and Christoffersen 2021) allows flexible modeling on the cumulative hazard scale, in the same spirit as the Royston-Parmar model (Royston and Parmar 2002), using either a fully parametric or a penalized approach. Excess hazard models can be fitted and clustered data can be accounted for by the inclusion of a gamma or log-normally distributed frailty.

Regarding other existing R packages allowing the inclusion of random effects to analyze correlated survival times, **frailtypack** is probably the most developed (Rondeau, Marzroui, and Gonzalez 2012), with functions for fitting shared and nested frailty models as well as joint modeling of multiple time-to-event processes. Users can specify either a gamma or a log-normal frailty distribution, and estimated parameters are obtained by using a penalized likelihood framework. In the full likelihood framework, the **parfm** package has been developed for shared frailty models with a parametric distribution associated to the time-to-event (Munda, Rotolo, and Legrand 2012), with many choices supported for the frailty distributions and parametric baseline hazards. Other R packages developed for fitting random-effect models on time-to-event data include **survival** (Therneau 2021) (via the `frailty()` element that can be added to the formula of `survreg()`) for parametric survival regression model, and, for semi-parametric hazard models, the **coxme** (Therneau 2020) and **frailtyEM** (Balan and Putter 2019) packages, among others. However, these packages do not offer functionalities for excess hazard regression modeling.

The **mexhaz** package (Charvat and Belot 2021), available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=mexhaz, allows for both flexible specification of the hazard and shared frailty modeling in the excess hazard setting, thus complementing the existing R packages. Besides the Weibull model, models using piecewise constant functions or splines (B-splines and restricted cubic splines) to describe the logarithm of the hazard are implemented. Time-dependent effects of covariates and delayed entry times can be accounted for and clustered data can be modeled through the inclusion of a log-normally distributed frailty. Table 1 summarizes the functionalities offered by **mexhaz** compared to some of the packages already cited.

As regards the domain of application, the **rstpm2** package is the one that most closely matches **mexhaz**'s capabilities. However, while **rstpm2** proposes a flexible modeling on the cumulative hazard scale, **mexhaz** offers modeling on the hazard scale. These equally valid approaches

Package	**mexhaz**	**rstpm2**	**flexsurv**	**frailtypack**	**parfm**	**survival**
Function	`mexhaz`	`stpm2; pstpm2`	`flexsurvreg;` `flexsurvspline`	`frailtyPenal`	`parfm`	`coxph;` `survreg`
Parametric models	Exponential, Weibull	Weibull, generalized gamma	Built-in or user-defined distributions	Exponential, Weibull	Exponential, Weibull, Gompertz, log-normal, log-logistic	Built-in or user-defined distributions
Flexible modeling of the (log-)hazard	Log-hazard scale: piecewise constant, B-splines (up to cubic), natural cubic splines	-	-	Hazard scale: penalized cubic M-splines	-	-
Flexible modeling of the log-cumulative hazard	-	Smoothed functions implemented in R (e.g., natural cubic splines)	Natural cubic splines	-	-	-
Position of spline knots	User-defined	Quantile of the event distribution; user-defined	Quantile of the event distribution; user-defined	Quantile of the event distribution; equispaced knots	-	-
Time-dependent effects	Yes	Yes	Yes	Yes	No	Yes
Shared frailty model for left-truncated data	Yes	No	No	Yes	Yes	No
Shared frailty distribution	Log-normal	Gamma, log-normal	-	Gamma, log-normal	Gamma, log-normal, positive stable, inverse Gaussian	Gamma, log-normal, log-t
Excess hazard modeling	Yes	Yes	Yes	No	No	No
Available outputs						
Prediction of the hazard / survival function	Yes	Yes	Yes	Yes	No	Yes
Gradient of the predicted values	Yes	No	No	No	No	No
Empirical Bayes estimates	Yes	No	-	Yes (with their variance only for Gamma frailty)	Yes (but not their variance)	Yes

Table 1: Comparison of the functionalities provided by **mexhaz** and other R packages for hazard regression modeling with excess hazard or shared frailty.

have advantages and limitations: modeling on the cumulative hazard scale is usually faster because there is no need for integration of the hazard; on the other hand, it might suffer from problems due to the instability of numerical differentiation. Besides, the interpretation of results from models defined on the cumulative hazard scale can present difficulties when multiple time-dependent effects are included (see Section 7.6.3 in Royston and Lambert 2011). It should also be noted that different modeling choices result in different constraints: models of the logarithm of the cumulative excess hazard (as in **rstpm2**) impose a constraint of positivity for the overall hazard, so that the excess hazard can become negative, while modeling the logarithm of the excess hazard, as is performed in **mexhaz**, imposes positivity of the excess hazard.

The aim of this paper is to present the full likelihood-based approach implemented in **mexhaz** for fitting flexible regression models defined on the hazard scale. The package allows the user to deal with i) time-dependent effects of covariates, ii) correlated survival times and iii) estimation of the disease-specific mortality hazard without relying on cause of death information (excess hazard model). In the first section, we present the general framework of flexible hazard-based regression models and describe their extension to excess hazard and mixed-effect (possibly, excess) hazard models. We then proceed to illustrate the use of the package in the second section and finally compare the results obtained in **mexhaz** with those of other packages offering similar functionalities.

2. Flexible parametric hazard-based regression model

2.1. General framework

In the following, we are concerned with time-to-event data, i.e., data recording the occurrence of an event (death, disease, relapse, etc.) along time. In this context, "survival at time t" refers to the state of not having presented the event by time t. The observed survival time, t, of an individual can then be seen as the realization of a non-negative random variable T. If we denote by f the probability density function of T, the cumulative distribution function, F, is defined by the relationship:

$$F(t) = P(T \leq t) = \int_0^t f(u) \, \mathrm{d}u$$

The survival at time t, representing the probability of being "alive" (free of the event) at t, is then defined as $S(t) = 1 - F(t)$.

Besides, a central quantity in survival analysis is the hazard, λ, representing the instantaneous failure rate (expressed as a number of events per person-time) and defined formally as:

$$\lambda(t) = \lim_{\mathrm{d}t \to 0} \frac{P(t \leq T < t + \mathrm{d}t \mid T \geq t)}{\mathrm{d}t}$$

The hazard is linked to f and S through the following relationship:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

from which we can derive:

$$S(t) = \exp\left\{-\int_0^t \lambda(u) \, \mathrm{d}u\right\}$$

Due to the presence of right censoring and left truncation, most methods developed for the analysis of time-to-event data are based on the hazard (Cox and Oakes 1984; Geskus 2015), and we will focus in this paper on hazard-based regression models.

Derivation of the likelihood

First of all, let us define some notations. For each individual j, $j = 1, \ldots, n$, let t_{0j} denote the time at entry, t_j the observed failure time (which is defined as the minimum between the survival time and the censoring time), and δ_j an indicator variable taking the value 1 in case of occurrence of the event at t_j and 0 in case of censoring.

Suppose that we want to describe the hazard in the study population by a function of time and of a vector of covariates (such as age, gender, etc.) parameterized by $\boldsymbol{\beta}$, the vector of parameters to be estimated. Under the hypothesis of non-informative censoring, the contribution of the unknown censoring distribution can be omitted from the likelihood function and we can thus write the likelihood for individual j with covariates $\mathbf{x}_j$ as (Kalbfleisch and Prentice 2002):

$$
\begin{aligned}
\mathrm{L}_j(\boldsymbol{\beta}) &= \frac{f(t_j, \mathbf{x}_j)^{\delta_j} S(t_j, \mathbf{x}_j)^{1-\delta_j}}{S(t_{0j}, \mathbf{x}_j)} \\
&= \lambda(t_j, \mathbf{x}_j)^{\delta_j} \exp\left\{ -\int_{t_{0j}}^{t_j} \lambda(u, \mathbf{x}_j)\, \mathrm{d}u \right\}
\end{aligned}
$$

Thus, the log-likelihood function for the whole population is obtained as the sum of the individual contributions:

$$
\begin{aligned}
\ell(\boldsymbol{\beta}) &= \sum_{j=1}^{n} \log\bigl(\mathrm{L}_j(\boldsymbol{\beta})\bigr) \\
&= \sum_{j=1}^{n} \left\{ \delta_j \log\bigl(\lambda(t_j, \mathbf{x}_j)\bigr) - \int_{t_{0j}}^{t_j} \lambda(u, \mathbf{x}_j)\, \mathrm{d}u \right\}
\end{aligned} \tag{1}
$$

Note that if $t_{0j} = 0$ for all individuals (no late entries), the log-likelihood can be written:

$$
\ell(\boldsymbol{\beta}) = \sum_{j=1}^{n} \left\{ \delta_j \log\bigl(\lambda(t_j, \mathbf{x}_j)\bigr) - \Lambda(t_j, \mathbf{x}_j) \right\}
$$

where $\Lambda(t) = \int_0^t \lambda(u)\, \mathrm{d}u$ is the cumulative hazard.

2.2. Extension to net survival analysis: Excess hazard regression model

In the context of the analysis of the mortality of cancer patients, we usually want to take into account the following situation: cancer patients may die from their cancer (either directly as a consequence of the natural progression of the disease, or indirectly as a consequence of the treatment) but they might also die from other causes like individuals sharing the same characteristics (age, gender, birth cohort, etc.) but not diagnosed with cancer. In other words, we would like to quantify the excess mortality that cancer patients experience because of their disease. One possible way to answer this question is to use the so-called excess hazard approach: it is based on the idea that the overall hazard may be decomposed into a

sum of two hazards, one taking into account the excess mortality that can be attributed to the disease under study, λ_e, and the other that takes care of all the other possible causes of death, λ_{oth}. The objective is then to estimate λ_e while considering λ_{oth} as known. Provided that the prevalence of the disease under study (or at least, that the mortality from this disease) is reasonably low in the general population, λ_{oth} can be approximated by the population (or expected) mortality hazard, λ_p, usually obtained from national statistics institutes and described as a function of demographic variables, such as sex, age, year, etc. This results in the following model:

$$\lambda(t, \mathbf{x}, \tilde{\mathbf{z}}) = \lambda_e(t, \mathbf{x}) + \lambda_p(a + t, y + t, \tilde{\mathbf{z}}) \tag{2}$$

where a and y represent age at diagnosis and year of diagnosis, respectively, $\mathbf{x}$ is a vector of covariates and $\tilde{\mathbf{z}}$ a vector of demographic variables used to define the population mortality ($\tilde{\mathbf{z}}$ is usually a subset of $\mathbf{x}$) excluding age and year of diagnosis. In practice, λ_p is obtained from population mortality tables and depends on age, year of diagnosis and other variables such as gender, county of residence (Estève *et al.* 1990; Remontet *et al.* 2007; Pohar Perme *et al.* 2012). In the following, we define $\mathbf{z} = (a, y, \tilde{\mathbf{z}})$ and write conveniently λ_p as a function of t and $\mathbf{z}$.

The excess hazard model consists in modeling λ_e by a function of time and covariates parameterized by $\boldsymbol{\beta}$. The log-likelihood is obtained by replacing λ in Equation 1 by its expression under the excess hazard assumption (Equation 2):

$$\ell(\boldsymbol{\beta}) = \sum_{j=1}^{n} \left\{ \delta_j \log\left(\lambda_e(t_j, \mathbf{x}_j) + \lambda_p(t_j, \mathbf{z}_j)\right) - \int_{t_{0j}}^{t_j} \left(\lambda_e(u, \mathbf{x}_j) + \lambda_p(u, \mathbf{z}_j)\right) \mathrm{d}u \right\} \tag{3}$$

The quantity $\sum_{j=1}^{n} \int_{t_{0j}}^{t_j} \lambda_p(u, \mathbf{z}_j) \, \mathrm{d}u$ does not depend on the parameters to be estimated and is usually dropped when maximizing the likelihood. This implies that only the value of the population hazard λ_p at the end of follow-up for each individual is necessary to specify the likelihood.

Also note that if we set λ_p to 0 for all individuals, Equation 2 simplifies to Equation 1 and specifies a model for the overall hazard.

2.3. Extension to hierarchical survival data: Shared frailty model

Now suppose that individuals come from different clusters (e.g., geographical areas): we thus have a two-level hierarchical structure with individuals nested in clusters. People from the same cluster share common characteristics so that their survival times are correlated. In order to take this structure into account, one possibility is to extend our previous model by including a random effect at the cluster level.

First of all, let us refine our notations. For each individual j, $j = 1, \ldots, n_i$ from cluster i, $i = 1, \ldots, D$, let t_{ij} denote the observed failure time and δ_{ij} an indicator variable taking the value 1 in case of an event and 0 in case of censoring. Our mixed-effect excess hazard regression model can be written as:

$$\lambda(t, \mathbf{x}, \mathbf{z}, w) = \lambda_{me}(t, \mathbf{x}, w) + \lambda_p(t, \mathbf{z})$$

with

$$\lambda_{me}(t, \mathbf{x}, w) = \lambda_e(t, \mathbf{x}) \exp(w)$$

where w is a random effect assumed to be normally distributed with mean 0 and variance σ^2. Note that this model can be parameterized in terms of $u = \exp\{w\}$, a quantity known as the shared frailty ("shared frailty models" is an equivalent term for designing mixed-effect hazard models): the distributional assumption of our model thus corresponds to a log-normally distributed shared frailty.

In the case of non-left truncated data (i.e., $t_{0ij} = 0$ for all individuals), the likelihood for a single observation j from cluster i conditional on the value of the random effect is:

$$\mathrm{L}_{ij}^C(\boldsymbol{\beta} \mid w_i) = \{\lambda_{me}(t_{ij}, \mathbf{x}_{ij}, w_i) + \lambda_p(t_{ij}, \mathbf{z}_{ij})\}^{\delta_{ij}} S(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}, w_i)$$

with

$$S(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}, w_i) = \exp\{-\Lambda_{me}(t_{ij}, \mathbf{x}_{ij}, w_i) - \Lambda_p(t_{ij}, \mathbf{z}_{ij})\} \tag{4}$$

In practice, the last term of the exponential in Equation 4 is omitted from the estimation procedure because it has no impact on the maximization of the likelihood. The conditional likelihood for cluster i is:

$$\mathrm{L}_i^C(\boldsymbol{\beta} \mid w_i) = \prod_{j=1}^{n_i} \mathrm{L}_{ij}^C(\boldsymbol{\beta} \mid w_i)$$

Then, the marginal likelihood for cluster i is obtained by integrating the conditional likelihood over the distribution of the random effect:

$$\mathrm{L}_i^M(\boldsymbol{\beta}, \sigma) = \int_{-\infty}^{\infty} \mathrm{L}_i^C(\boldsymbol{\beta} \mid w_i)\, \phi(w_i, 0, \sigma)\, \mathrm{d}w_i \tag{5}$$

where $\phi(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$.

The model parameters $(\boldsymbol{\beta}^\top, \sigma)^\top$ can then be estimated by maximizing the full log-likelihood:

$$\ell(\boldsymbol{\beta}, \sigma) = \sum_{i=1}^{D} \log\{\mathrm{L}_i^M(\boldsymbol{\beta}, \sigma)\} \tag{6}$$

Note once again that we obtain the likelihood for a standard mixed-effect hazard model when λ_p is set to 0 (which corresponds to modeling the overall hazard).

In the presence of left truncation, the formulation of the likelihood must take account of the fact that the delayed entry times are conditional on the value of the frailty (defined at $t = 0$). Indeed, individuals with a lower value of the frailty are more likely to be alive up to a given time, so that the distribution of the frailty in a population with delayed entry times differs from the distribution of the frailty in the original population (Wienke 2010; Van den Berg and Drepper 2016; Crowther, Andersson, Lambert, Abrams, and Humphreys 2016). The marginal likelihood for cluster i then becomes:

$$\mathrm{L}_i^M(\boldsymbol{\beta}, \sigma) = \frac{\displaystyle\int_{-\infty}^{\infty} \mathrm{L}_i^C(\boldsymbol{\beta} \mid w_i)\, \phi(w_i, 0, \sigma)\, \mathrm{d}w_i}{\displaystyle\int_{-\infty}^{\infty} \prod_{j=1}^{n_i} \left\{S(t_{0ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}, w_i)\right\} \phi(w_i, 0, \sigma)\, \mathrm{d}w_i} \tag{7}$$

where t_{0ij} is the delayed entry time for individual j from cluster i. Also note that, here again, when modeling the excess mortality hazard, the denominator involves a term related to the expected hazard that can be omitted from the likelihood maximization procedure.

2.4. Model specificities

Functional forms for the baseline hazard and the time-dependent effects

In **mexhaz**, the hazard, λ (in the classical setting), or the excess hazard, λ_e (in the excess hazard setting), is modeled as a function of time and some covariates depending on a vector of parameters $\boldsymbol{\beta}$. In the following, we will use the generic notation $\widetilde{\lambda}$ to refer to either λ or λ_e depending on the setting.

The user can choose between the so-called Weibull model and flexible parametric models of the hazard based on either a piecewise constant function, B-splines (up to degree 3) or restricted cubic splines.

Concerning the Weibull model, the general expression of the hazard, $\widetilde{\lambda}$, as a function of time t and covariates $\mathbf{x}$ (assuming that the first N variables are modeled with a proportional effect and the M following are modeled with a time-dependent effect), is:

$$\widetilde{\lambda}(t, \mathbf{x}) = \rho(\mathbf{x})\theta(\mathbf{x})t^{\theta(\mathbf{x})-1} \tag{8}$$

with ρ and θ respectively the scale and shape parameters of the Weibull function depending on $\mathbf{x}$ through the relationships:

$$\begin{cases} \rho(\mathbf{x}) = \exp\Big\{\gamma_0 + \sum_{k=1}^{M+N} \gamma_k\, x_k\Big\} \\ \theta(\mathbf{x}) = \exp\Big\{\xi_0 + \sum_{m=N+1}^{M+N} \xi_m\, x_m\Big\} \end{cases} \tag{9}$$

where

- γ_0 is the logarithm of the constant scale parameter,

- the γ_k's ($k \in \{1, \ldots, N\}$) are the coefficients corresponding to the variables modeled with a proportional effect,

- the γ_k's ($k \in \{N + 1, \ldots, M + N\}$) are the coefficients corresponding to the non-time dependent part of the effect of the variables modeled with a time-dependent effect,

- ξ_0 is the logarithm of the constant shape parameter,

- the ξ_m's ($m \in \{N + 1, \ldots, M + N\}$) are the coefficients corresponding to the time-dependent part of the effect of the variables modeled with a time-dependent effect.

For the piecewise constant- and spline-based models, the user is free to choose the number of knots and specify their locations. The time-dependent effects of covariates are parametrized as interaction terms between the covariates and the baseline hazard, thus leading to the same functional form for the time-dependent effect than the one used for the baseline hazard. So for example, if the baseline hazard is modeled using a quadratic B-spline with one knot at 1 year, then the time-dependent effect of a covariate will be parametrized using a quadratic B-spline with one knot at 1 year.

With the same conventions as for the Weibull model, the general expression of the hazard is now:

$$\widetilde{\lambda}(t,\mathbf{x}) = \exp\Big\{\underbrace{\gamma_0 + \sum_{k=1}^{M+N}\gamma_k\,x_k}_{\text{Time-fixed part}} + \underbrace{\sum_{l=1}^{L}\Big(\xi_{l0} + \sum_{m=N+1}^{M+N}\xi_{lm}\,x_m\Big)\text{FT}_l(t)}_{\text{Time-dependent part}}\Big\} \tag{10}$$

where

- FT_l are the basis functions of time used to describe the baseline hazard and the time-dependent effects of covariates. In practice, **mexhaz** allows the use of i) B-splines of degree 1 to 3 (in which case L is the sum of the degree of the spline and of the number of interior knots), ii) restricted cubic B-splines (in which case L is equal to 1 plus the number of interior knots), and iii) piecewise constant functions, in which case L is equal to 1 plus the number of interior knots;

- γ_0 is the coefficient corresponding to the constant term (or "intercept") of the model,

- the γ_k's ($k \in \{1,\dots,N\}$) are the coefficients corresponding to the variables modeled with a proportional effect,

- the γ_k's ($k \in \{N+1,\dots,M+N\}$) are the coefficients corresponding to the non-time dependent part of the effect of the variables modeled with a time-dependent effect,

- the ξ_{l0}'s are the coefficients corresponding to the spline modeling the logarithm of the baseline hazard,

- the ξ_{lm}'s ($m > N$) are the coefficients corresponding to the modeling of the time-dependent effect of the variables (obtained by considering interaction terms with the function used to model the baseline hazard).

For example, for a model in which the baseline hazard is described with a quadratic B-spline with two knots (i.e., requiring four basis functions, named here $\text{BS}_1,\dots,\text{BS}_4$, in addition to the intercept), with the variable x_1 modeled with a proportional (i.e, constant in time) effect, and the variable x_2 modeled with a time-dependent effect, Equation 10 becomes:

$$\begin{aligned}
\widetilde{\lambda}(t,x_1,x_2) &= \exp\Big\{\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \sum_{l=1}^{4}\xi_{l0}\,\text{BS}_l(t) + x_2\sum_{l=1}^{4}\xi_{l2}\,\text{BS}_l(t)\Big\}\\
&= \underbrace{\exp\Big\{\gamma_0 + \sum_{l=1}^{4}\xi_{l0}\,\text{BS}_l(t)\Big\}}_{\lambda_0(t)}\underbrace{\exp\Big\{\gamma_1 x_1 + \big(\gamma_2 + \sum_{l=1}^{4}\xi_{l2}\,\text{BS}_l(t)\big)x_2\Big\}}_{f(t)}
\end{aligned}$$

That is, the hazard can be expressed in the form:

$$\widetilde{\lambda}(t,x_1,x_2) = \lambda_0(t)\exp(\gamma_1 x_1 + f(t)x_2)$$

with the restriction that f be based on the same basis functions of time as the ones used to model the logarithm of the baseline hazard.

Calculation of the cumulative hazard

Computation of the log-likelihood requires the calculation of the cumulative hazard (see Equations 1, 3 or 4). Depending on the choice of function used to describe the hazard, the `mexhaz()` function computes the cumulative hazard in different ways:

- for the Weibull model, or when the logarithm of the hazard is described by a piecewise constant function or a B-spline of degree 1, the calculation is based on the analytical formula for the cumulative hazard;

- when the log-hazard is described by a quadratic or cubic B-spline or by a restricted cubic spline, the cumulative hazard is obtained by numerical integration, more precisely by Gauss-Legendre quadrature.

The Gauss-Legendre quadrature is a numerical integration technique which approximates the integral of a function defined on $[-1, 1]$ by a weighted sum using G pre-specified weights and nodes (Mathews and Fink 1999). The G-point Gauss-Legendre rule is exact for polynomials functions of degree less or equal than $2G - 1$. By applying a simple change of variable, the Gauss-Legendre quadrature rule can be used for approximating the integral of functions defined on the interval $[t_0, t_1]$ according to the general formula:

$$\int_{t_0}^{t_1} \widetilde{\lambda}(u)\,\mathrm{d}u \approx \frac{t_1 - t_0}{2} \sum_{g=1}^{G} w_g \, \widetilde{\lambda}\Big(\frac{t_0 + t_1}{2} + \frac{t_1 - t_0}{2}\, z_g\Big)$$

where w_g and z_g are the weights and nodes for the G-point GL rule, respectively. Those G weights and nodes are available in the **statmod** R package (Giner and Smyth 2016).

In the `mexhaz()` function, we apply the Gauss-Legendre quadrature rule on subintervals defined by the interior knots used to define the spline bases. More precisely, to integrate the hazard on $[t_0, t_1]$, supposing that this interval contains K of the knots used to define the spline and renumbering for convenience these knots so that $k_0 = t_0, \ldots, k_{K+1} = t_1$, we write:

$$\int_{t_0}^{t_1} \widetilde{\lambda}(u)\,\mathrm{d}u = \sum_{j=0}^{K} \int_{k_j}^{k_{j+1}} \widetilde{\lambda}(u)\,\mathrm{d}u$$

and apply the G-point Gauss-Legendre quadrature rule to each $\int_{k_j}^{k_{j+1}} \widetilde{\lambda}(u)\,\mathrm{d}u$. By default in `mexhaz()`, G is set to 20.

Computation of the marginal cluster-specific likelihoods

The marginal likelihood L_i^M in Equation 5 as well as the numerator and denominator of Equation 7 do not have a closed analytical form. It is thus necessary to use numerical methods to approximate their value.

The Gauss-Hermite quadrature is a numerical integration technique that allows the evaluation of integrals of the form:

$$\int_{-\infty}^{\infty} f(x)\, \exp(-x^2)\,\mathrm{d}x$$

by computing a weighted sum of the function f evaluated at particular points called the quadrature nodes:

$$\int_{-\infty}^{\infty} f(x) \exp(-x^2) \, \mathrm{d}x \approx \sum_{q=1}^{Q} \rho_q^H \, f(x_q^H)$$

The Q nodes x_q^H and weights ρ_q^H are computed from the zeros of the Q-th order Hermite polynomial and in the context of our work, were obtained through the use of the R package **statmod** which makes use of an algorithm previously developed by Golub and Welsch (1969).

With a simple transformation of the weights and nodes, the same principle can be used to evaluate integrals of the form:

$$\int_{-\infty}^{\infty} f(x) \, \phi(x, \mu, \sigma) \, \mathrm{d}x \approx \sum_{q=1}^{Q} \rho_q^N \, f(x_q^N(\mu, \sigma)) \tag{11}$$

The transformed nodes, x_q^N, are now functions of μ and σ, respectively the mean and the standard deviation of the normal density function ϕ. In particular, these nodes and weights do not depend on the function f appearing in the integrand. This means that the positions of the nodes might not cover adequately the region of variation of the function resulting in i) a poor approximation of the integral and ii) the necessity of using a large number of nodes to try to improve that approximation.

Note also that, by defining the function $g(x, \mu, \sigma) = f(x)\phi(x, \mu, \sigma)$, we can rewrite Equation 11 as follows:

$$\int_{-\infty}^{\infty} g(x, \mu, \sigma) \, \mathrm{d}x \approx \sum_{q=1}^{Q} \rho_q^* \, g(x_q^*, \mu, \sigma)$$

with

$$\begin{cases} x_q^*(\mu, \sigma) = \mu + \sigma\sqrt{2} \, x_q^H \\ \rho_q^* = \dfrac{1}{\sqrt{\pi}} \exp\{(x_q^H)^2\} \, \rho_q^H \end{cases}$$

This last formulation is seldom presented but it makes the comparison with the adaptive Gauss-Hermite quadrature more straightforward.

The idea of the adaptive Gauss-Hermite quadrature (Liu and Pierce 1994; Pinheiro and Bates 1995) is to transform the integrand in order to obtain a new quadrature formula in which the nodes and corresponding weights depend on the function f: the nodes are translated and rescaled so that they cover the region where the integrand varies most, that is, around its mode. These specific nodes and weights depend on the location and the shape of the integrand of Equation 5, and are defined by using a Laplace approximation. More precisely, the adaptive Gauss-Hermite quadrature method requires, for each cluster, the computation of the mode of the integrand, μ_i, and σ_i defined as the negative of the inverse of the second derivative of the logarithm of the integrand evaluated at μ_i. These values are then used to transform the nodes and weights according to the following relationships:

$$\begin{cases} x_q^A(\mu_i, \sigma_i) = \mu_i + \sigma_i\sqrt{2} \, x_q^H \\ \rho_q^A(\sigma_i) = \sigma_i\sqrt{2} \exp\{(x_q^H)^2\} \, \rho_q^H \end{cases}$$

This transformation results in a much better approximation of the integral with a small number of quadrature points at the cost of extra computational time because of the evaluation of the cluster-specific μ_i's and σ_i's, which requires calculations involving the first and second derivatives of the integrand.

When applying Gauss-Hermite quadrature to the approximation of the cluster-specific marginal likelihood, we obtain:

$$\mathrm{L}_i^M(\boldsymbol{\beta}, \sigma) \approx \sum_{q=1}^{Q} \rho_q^A(\sigma_i)\,\mathrm{L}_i^C(\boldsymbol{\beta}, \sigma \mid x_q^A(\mu_i, \sigma_i))\phi(x_q^A(\mu_i, \sigma_i), 0, \sigma)$$

In **mexhaz**, the cluster-specific μ_i's and σ_i's are estimated at each iteration of the optimization algorithm, and by default, the number of quadrature points Q is set to 10.

Optimization procedure

Once the full log-likelihood according to the context of the study (overall or excess mortality hazard, with or without a hierarchical structure) has been defined (as in Equations 1, 3, or 6), the `nlm` function is used by default in **mexhaz** to estimate the parameters $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma})^\top$. However, the user can alternatively choose the `optim` function, with all the different optimization algorithms proposed. The estimated covariance matrix, $\widehat{\Sigma}_{\boldsymbol{\theta}}$, is obtained as the inverse of the negative of the Hessian matrix, the standard errors of the parameters being the square root of the diagonal elements.

Because the Hessian matrix returned by the optimization algorithm might not be very accurate, the user can ask for a better approximation via the `numHess = TRUE` option of the `mexhaz()` function. In that case, the Hessian is evaluated through the function `hessian()` from the **numDeriv** R package (numerical derivation based on the Richardson method).

Shrinkage estimates

The cluster-specific random effects, commonly called "shrinkage estimates" or "empirical Bayes estimates", can be obtained as the modes of the integrand appearing in Equation 5 evaluated at the maximum likelihood value of the parameters:

$$\mu_i = \frac{1}{\widehat{\sigma}\sqrt{2\pi}}\mathrm{argmax}_w\Big(\mathrm{L}_i^C(\widehat{\boldsymbol{\beta}} \mid w)\,\exp\Big\{-\frac{w^2}{2\widehat{\sigma}^2}\Big\}\Big)$$

An approximate variance for these shrinkage estimates is obtained by the following formula (Booth and Hobert 1998):

$$\mathrm{Var}(\mu_i) \approx \sigma_i^2 + \Big(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\Big)^\top \widehat{\Sigma}_{\boldsymbol{\theta}}\Big(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\Big)$$

where σ_i is the inverse of minus the second derivative of the logarithm of the integrand appearing in Equation 5 evaluated at μ_i and $\Big(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\Big)$ is the gradient of μ_i, i.e., the vector of partial derivatives of μ_i relative to the model parameters, evaluated at the maximum likelihood value of these parameters.

The approximate covariances between μ_i and the fixed effect parameters $\boldsymbol{\beta}$ are given by:

$$\mathrm{Cov}(\mu_i, \boldsymbol{\beta}) \approx \Big(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}})\Big)^\top \widehat{\Sigma}_{\boldsymbol{\beta}}$$

And the approximate covariance between two cluster-specific random effects μ_i and μ_j is given by:

$$\mathrm{Cov}(\mu_i, \mu_j) \approx \left(\frac{\partial \mu_i}{\partial \boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\right)^\top \widehat{\Sigma}_{\boldsymbol{\theta}} \left(\frac{\partial \mu_j}{\partial \boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})\right)$$

Predictions and confidence intervals

In **mexhaz**, we provide tools to predict and plot the modeled hazard and the corresponding survival. In particular, in the excess hazard setting, these will correspond to the excess hazard and net survival if the population mortality rate is specified, and to the overall hazard and the overall survival if this population mortality rate is omitted from the model specification. To derive the corresponding confidence intervals, the user has the possibility to use either the delta method or a Monte-Carlo simulation-based method. For the Monte-Carlo method, the user can specify the number of simulations used.

3. Illustration

3.1. Introduction

The main function used for model fitting, `mexhaz()`, requires the following arguments:

- `formula`: a formula with the response on the left of the `~` operator, and the linear predictor on the right. The response must be of the form `Surv(time, event)`, following the classical survival model formulation popularized by the R package **survival**;

- `data`: the name of the dataset;

- `base`: the functional form that should be used to model the baseline hazard. Selection can be made between the following options: `"weibull"` for a Weibull hazard, `"exp.bs"` for a hazard described by the exponential of a *B*-spline (only *B*-splines of degree 1, 2 or 3 are accepted), `"exp.ns"` for a hazard described by the exponential of a restricted cubic spline (also called "natural spline"), `"pw.cst"` for a piecewise constant hazard;

- `degree`: specifies the degree of the *B*-spline;

- `knots`: if `base = "exp.bs"` or `"exp.ns"`, `knots` is the vector of interior knots of the spline. If `base = "pw.cst"`, `knots` is the vector defining the endpoints of the time intervals on which the hazard is assumed to be constant. By default, `knots = NULL` (that is, it produces a *B*-spline with no interior knots if `base = "exp.bs"`, a linear *B*-spline with no interior knots if `base = "exp.ns"`, or a constant hazard over the whole follow-up period if `base = "pw.cst"`);

- `expect`: name of the variable (from the dataset) defining the expected hazard (for excess hazard model estimation). By default `expect = NULL`, corresponding to a standard hazard regression model (which is a model for the overall hazard);

- `random`: name of the variable defining the cluster membership (for mixed effect hazard model estimation). By default `random = NULL`, corresponding to a fixed effect survival model.

The reader is referred to the help page of the function for more details on the arguments of `mexhaz()`.

Two simulated datasets (Charvat *et al.* 2016) provided as part of our package are used in this section in order to illustrate the various functionalities of the `mexhaz()` function.

The `simdatn1` dataset has 4000 rows and 8 columns and contains the following variables:

- `age`: Age at diagnosis (continuous);

- `agecr`: Centered and rescaled age variable corresponding to $(age - 70)/100$;

- `depindex`: Deprivation index (continuous) defined at the cluster level;

- `IsexH`: Gender $(0 = $ Female, $1 = $ Male$)$;

- `clust`: ID number of the cluster;

- `vstat`: Vital status $(0 = $ Censored, $1 = $ Dead$)$;

- `timesurv`: Follow-up time (year), administratively censored after 10 years;

- `popmrate`: Population (expected) mortality rate at the time of censoring or death. This was based on French population mortality tables depending on sex, age and calendar year.

The `simdatn2` dataset has the same structure and contains the same variables but data have been generated so that gender has a time-dependent effect.

For convenience, we define a new variable named `agec`, corresponding to the age centered around 70 years.

```
R> data("simdatn1", package = "mexhaz")
R> data("simdatn2", package = "mexhaz")
R> simdatn1$agec <- simdatn1$age - 70
R> simdatn2$agec <- simdatn2$age - 70
R> head(simdatn1, 3)

        age       agecr    depindex IsexH clust vstat   timesurv     popmrate
1 37.47094 -0.3252906   1.4995788     1    29     1 0.1650722 0.002166154
2 37.43171 -0.3256829  -0.7932424     1    39     1 0.3027221 0.002166154
3 38.23911 -0.3176089   1.2784658     1    17     1 0.2277427 0.002361141
       agec
1 -32.52906
2 -32.56829
3 -31.76089


R> head(simdatn2, 3)

        age       agecr    depindex IsexH clust vstat    timesurv     popmrate
1 30.15893 -0.3984107  -0.1930792     1    15     1 0.61613832 0.001472332
2 32.44655 -0.3755345   0.4983286     1    36     1 0.95682302 0.001639670
```

```
3 34.15436 -0.3584564  1.9553450     1    39     1 0.04132608 0.001730233
          agec
1 -39.84107
2 -37.55345
3 -35.84564
```

3.2. Flexible hazard-based regression models

Model fitting

Using the simulated dataset `simdatn1`, we show how to fit four models in which the overall mortality (i.e., without taking account of the population mortality rate) is described as a function of age at diagnosis (`agecr`) and gender (`IsexH`). The effect of these variables is assumed to be proportional (i.e., constant over time) but the baseline hazard is described by a different function in each model: Weibull hazard; piecewise-constant hazard with knots at 1, 2, 4, 6 and 8 years of follow-up; hazard described by the exponential of a cubic *B*-spline with knots at 1 and 5 years; and hazard described by the exponential of a restricted *B*-spline with knots at 1 and 5 years (in that case, the *B*-spline is constrained to be linear before 0 and after 10, which corresponds to the default boundary knots defined as 0 and the maximum of the observed follow-up times).

For convenience, we first define the model formula:

```
R> Form1 <- Surv(time = timesurv, event = vstat) ~ agec + IsexH
```

And then fit the four models described above:

```
R> ModWb <- mexhaz(formula = Form1, data = simdatn1, base = "weibull")
R> ModPw <- mexhaz(formula = Form1, data = simdatn1, base = "pw.cst",
+    knots = c(1, 2, 4, 6, 8))
R> ModBs <- mexhaz(formula = Form1, data = simdatn1, base = "exp.bs",
+    degree = 3, knots = c(1, 5))
R> ModNs <- mexhaz(formula = Form1, data = simdatn1, base = "exp.ns",
+    knots = c(1, 5))
```

We obtain the following results:

```
                  agec IsexH -2*log-lik N Param     AIC
Weibull          0.052 0.914  -6245.740       4 12499.48
Piecewise Cst    0.053 0.953  -6411.072       8 12838.15
Restricted Spline 0.052 0.924  -6329.249       6 12670.50
B-Spline         0.052 0.920  -6283.951       8 12583.90
```

Not surprisingly, the model which provides the best fit (according to the Akaike Information Criterion) is the Weibull model as the data were simulated using a Weibull hazard. Among the three other models, we note that the model in which the baseline hazard is described by the exponential of a cubic *B*-spline with two interior knots also provides a good fit. The question

of how to choose adequately the number and position of knots of spline functions is still open: the most frequently used methods consist in i) choosing the knots based on prior knowledge of the data-generating mechanism, or ii) determining the knots as the predefined percentiles of the distribution of the survival times of individuals who presented the event (Charvat *et al.* 2016). From an empirical point of view, the fit can also be checked a posteriori by comparing the average of the predicted survival curves in the study population with a non-parametric estimator of survival (or net survival in the excess hazard setting).

Prediction

The output of the `mexhaz()` function is an object of class 'mexhaz'. A `predict` method is defined for objects of class 'mexhaz': it allows the computation of hazard and survival estimates for a given time and a given vector of covariate values (see the help page of `predict.mexhaz()` for more details). More precisely, we can use `predict.mexhaz()` to predict both the survival and the corresponding hazard for i) several individuals with specific set of covariates at one pre-specified time, or ii) at several time points for one individual.

We illustrate these possibilities using the previously fitted models. The object `P.bs.10` corresponds to the prediction at 10 years (argument `time.pts`) for both female and male aged 70 years at diagnosis (`IsexH = c(0,1)` and `agec = 0`).

```
R> MyTime <- seq(0, 10, le = 1001)
R> MyData <- data.frame(agec = 0, IsexH = c(0, 1))
R> P.bs.10 <- predict(ModBs, time.pts = 10, data.val = MyData)
R> round(P.bs.10$results, 3)

  time.pts agec IsexH hazard hazard.inf hazard.sup  surv surv.inf surv.sup
1       10    0     0  0.089      0.061      0.131 0.243    0.224    0.262
2       10    0     1  0.223      0.152      0.329 0.029    0.023    0.035
```

The objects `P.bs0` and `P.bs1` correspond to predictions from `ModBs` for female (`Isex = 0`) and male (`Isex = 1`) aged 70 years at diagnosis from 0 to 10 years by increments of 0.01 years. They are used in the following section for graphical representation purposes.

```
R> P.bs0 <- predict(ModBs, time.pts = MyTime, data.val = MyData[1, ])
R> P.bs1 <- predict(ModBs, time.pts = MyTime, data.val = MyData[2, ])
```

By default, the confidence intervals are based on the Delta method, assuming the normality of the logarithm of the cumulative hazard. The function `predict.mexhaz()` allows the user to get the components of the gradient of the logarithm of the hazard and cumulative hazard through the `include.gradient = TRUE` argument. This might be useful for example if one is interested in estimating the confidence interval for a weighted sum of individual-specific survival curves.

Graphical results

The output of `predict.mexhaz()` is an object of class 'predMexhaz' that can be used by the functions `plot.predMexhaz()`, `lines.predMexhaz()` and `points.predMexhaz()` to plot the hazard and survival functions. By default, confidence intervals are also plotted (if present in
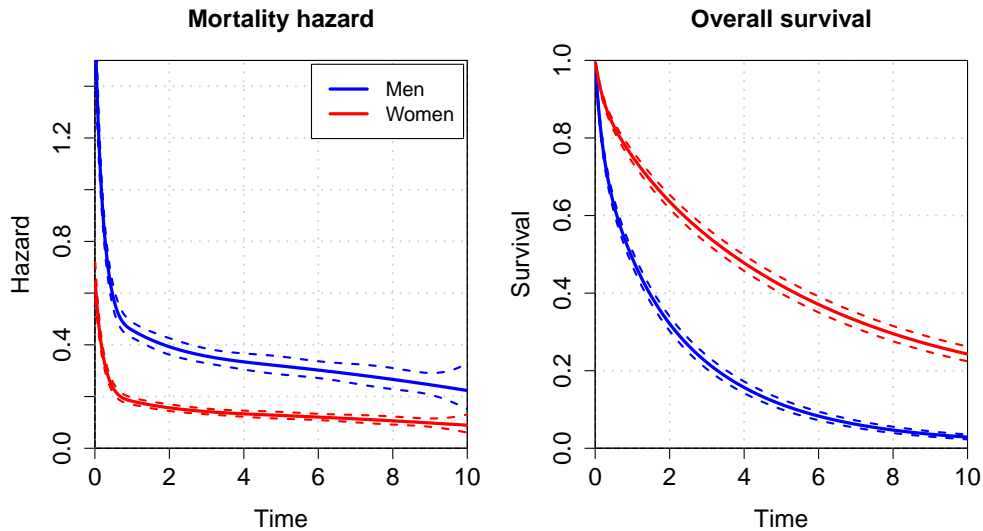
Figure 1: Mortality hazard and corresponding survival estimated for men and women aged 70.

the `predMexhaz` object). The following examples (corresponding to the graphical representation of the hazard and survival curves for men and women aged 70 based on the previously fitted models, see Figure 1), show how some of the standard arguments of the plotting functions can be used (see the help page of the package functions for more details).

```
R> plot(P.bs1, which = "hazard", ylim = c(0, 1.5), lwd = 2.5, col = "blue",
+    main = "Mortality hazard")
R> lines(P.bs0, which = "hazard", lwd = 2.5, col = "red")
R> plot(P.bs1, which = "surv", ylim = c(0, 1), lwd = 2.5, col = "blue",
+    main = "Overall survival")
R> lines(P.bs0, which = "surv", lwd = 2.5, col = "red")
```

In Figure 2, we represent the hazard and survival functions for women aged 70 years based on three of the previously fitted models. We can see that, although the hazards seem different, the survival estimates (as well as their confidence intervals, not shown here) are very similar.

Time-dependent effect

One of most commonly used assumption in survival analysis, mainly due to the popularity of the Cox model, is the proportionality of the hazards obtained for different values of the covariates. In other terms, the effects of the covariates, measured by the hazard ratio, is assumed to be constant over time, and the hazard for a specific value of a covariate is obtained by multiplying the baseline hazard by this constant. Although this assumption makes sense in many situations and greatly simplifies the estimation and interpretation of hazard-based survival models, there is sometimes no reason to think that the effect of a covariate should be constant over time.

The modelization of non-proportional effects of covariates is possible with the `mexhaz()` function, through the use of the `nph()` option in the model formula, and a graphical example is
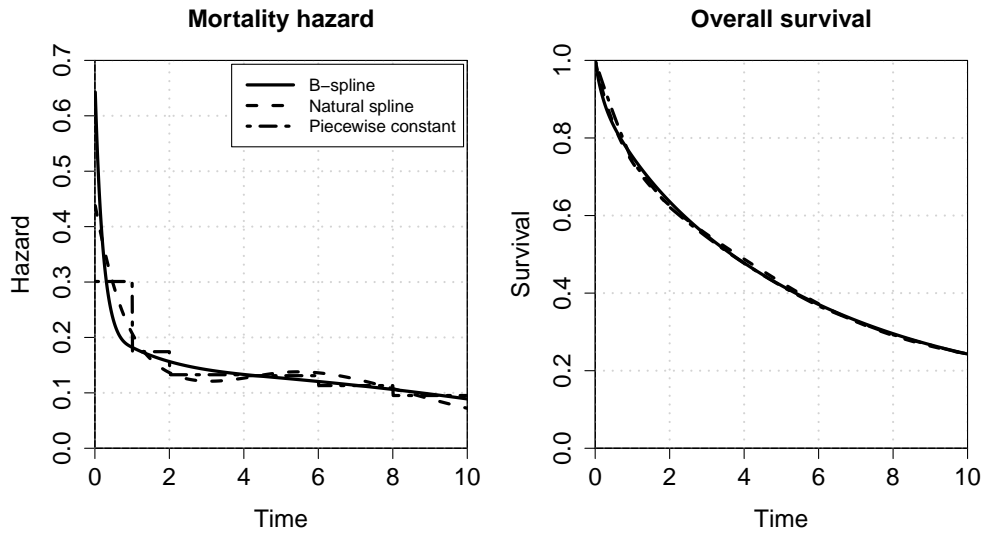
Figure 2: Mortality hazard and corresponding survival estimated by three different models.
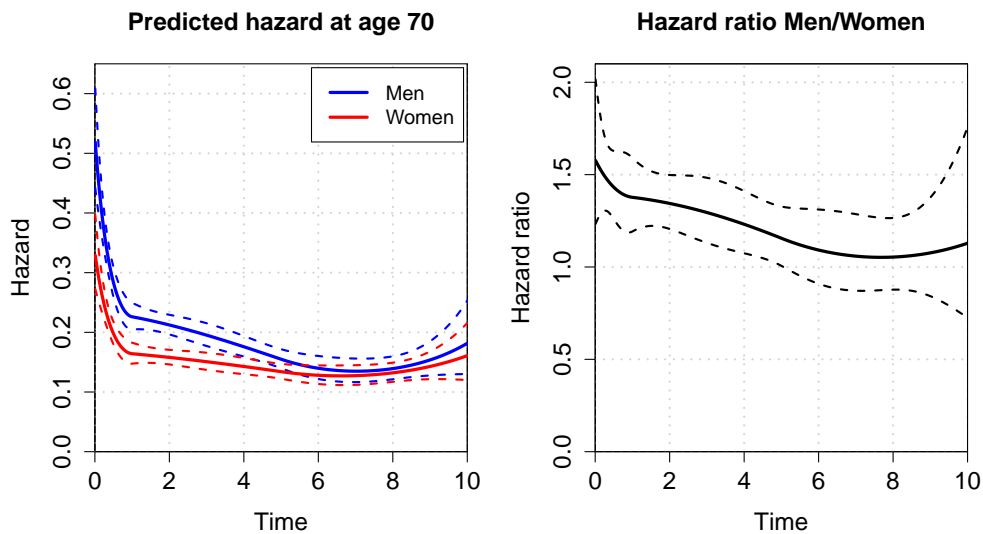

Figure 3: Example of time-dependent effect of gender.

shown in Figure 3. In the current state of development of the package, non-proportional effects are modeled as interaction terms between the covariates and the baseline hazard. It is planned in future versions to allow for more flexibility through the specification of different time functions for each non-proportional effect.

```
R> ModBs2.Nph <- mexhaz(Surv(time = timesurv, event = vstat) ~ agec + IsexH +
+    nph(IsexH), data = simdatn2, base = "exp.bs",
+    degree = 2, knots = c(1, 5))
```

Note on the modeling of non-linear effects of variables

For the purpose of illustrating the general syntax of the `mexhaz()` function, in this section and the following, we have modeled the effect of the continuous variable `agec` as linear (on the logarithm of the hazard). In real application, care should of course be taken about how to model the effect of such a variable by introducing, if necessary, non-linear terms. This can be done, as with most other R functions for regression models, by either i) creating new variables in the dataset corresponding to non-linear functions of the variable of interest, ii) using the `I()` operator within the model formula to generate these variables directly during model fitting, or iii) using functions such as `bs()` or `ns()` directly within the model formula.

3.3. Excess hazard regression model

General syntax

As we saw previously, the excess hazard regression approach to net survival estimation requires only the knowledge of the population mortality rate at the end of follow-up for each individual (in order to specify the likelihood). This is in contrast with non-parametric methods (such as the Pohar-Perme estimator) that require information on the population hazard at each event time.

As a consequence, unlike functions implementing non-parametric methods (such as the function `rs.surv()` from the **relsurv** package) for which the full lifetable has to be provided in the form of a `ratetable` object, the `mexhaz()` function only requires an extra variable (which has to be included in the dataset) corresponding to the population mortality rate at the end of follow-up.

The syntax for fitting an excess mortality hazard model with the `mexhaz()` function is thus simply obtained by specifying the name of the dataset variable containing the expected mortality rate through the `expected` argument:

```
R> ModBsExc <- mexhaz(Surv(timesurv, vstat) ~ agec + IsexH +
+    nph(IsexH), data = simdatn1, base = "exp.bs",
+    degree = 3, knots = c(1, 5), expected = "popmrate")

R> summary(ModBsExc)

Call:
mexhaz(formula = Surv(timesurv, vstat) ~ agec + IsexH + nph(IsexH),
    data = simdatn1, expected = "popmrate", base = "exp.bs",
    degree = 3, knots = c(1, 5))

Coefficients:
            Estimate    StdErr t.value    p.value
Intercept  -0.4329343 0.1028774 -4.2083 2.630e-05 ***
BS3.1      -1.3962071 0.1871848 -7.4590 1.064e-13 ***
BS3.2      -1.6915795 0.1839224 -9.1972 < 2.2e-16 ***
BS3.3      -1.6158372 0.3494209 -4.6243 3.877e-06 ***
BS3.4      -2.5209196 0.3721655 -6.7737 1.440e-11 ***
```

```
BS3.5        -2.5145466  0.4139937 -6.0739 1.365e-09 ***
agec          0.0453846  0.0018422 24.6364 < 2.2e-16 ***
IsexH         0.9680385  0.1215462  7.9644 2.148e-15 ***
IsexH*BS3.1   0.2032594  0.2272727  0.8943    0.3712
IsexH*BS3.2   0.0530386  0.2474255  0.2144    0.8303
IsexH*BS3.3  -0.7926945  0.5140445 -1.5421    0.1231
IsexH*BS3.4   0.6032433  0.5599749  1.0773    0.2814
IsexH*BS3.5  -0.0896992  0.6223925 -0.1441    0.8854
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Hazard ratios (for proportional effect variables):
        Coef     HR CI.lower CI.upper
agec  0.0454 1.0464   1.0427   1.0502
IsexH 0.9680 2.6328   2.0745   3.3412


log-likelihood: -5812.5676 (for 13 degree(s) of freedom)


number of observations: 4000, number of events: 3237
```

The parameter estimates now pertains to the effect of the covariates on the excess mortality hazard that patients are subject to because of their disease: from these parameters, excess hazard ratios can be calculated.

A note on the population mortality rate variable

In the previous example, the population mortality rate variable `popmrate` was already provided as part of the example dataset. However, it is usually necessary to create this variable from population mortality tables. We show here how this can be done using data from the **rstpm2** package, namely the `colon` dataset that contains $15,564$ observations on colon cancer patients, and the `popmort` dataset that provides the corresponding population mortality rates.

```
R> data("colon", package = "rstpm2")
R> data("popmort", package = "rstpm2")
R> head(colon, 3)


    sex age      stage mmdx yydx surv_mm surv_yy        status
1 Female  77    Distant    9 1977    16.5     1.5 Dead: cancer
2 Female  78 Localised   10 1978    82.5     6.5  Dead: other
3   Male  78    Distant   12 1978     1.5     0.5 Dead: cancer
                subsite       year8594 agegrp         dx       exit
1            Transverse Diagnosed 75-84    75+ 1977-09-07 1979-01-22
2   Coecum and ascending Diagnosed 75-84    75+ 1978-10-07 1985-08-22
3 Descending and sigmoid Diagnosed 75-84    75+ 1978-12-07 1979-01-22


R> head(popmort, 3)
```

```
  sex    prob        rate age year
1   1 0.96429 0.036363177   0 1951
2   1 0.99639 0.003616547   1 1951
3   1 0.99783 0.002172384   2 1951
```

The general principle is to compute from the available variables the age and year reached at the end of follow-up for each individual and then retrieve from the mortality table the value of the mortality rate corresponding to each individual-specific sex, age at exit and year at exit. It should be noted that the variables available (e.g., complete birth date versus age given as an integer) and the extension of the lifetables (e.g., ages are available until 99 years whereas some individuals in the dataset are 100 years or older at the end of follow-up) might introduce some differences in the selection of the appropriate mortality rate. For example, for an individual diagnosed in December 2015 at age 40 years and 8 months and censored after 6 months (i.e., in May 2016 at age 41 years and 2 months), the correct population mortality rate at end of follow-up corresponds to the mortality rate for year 2016 and age 41. However, if age and year of diagnosis are only available as integer (i.e., the diagnosis is made in 2015 at age 40), we can only compute an approximate age at exit of 40.5 years and a year at exit of 2015.5, which in terms of attained age and year at exit, corresponds to 40 years and 2015, thus leading to the selection of a different population mortality rate.

Because our objective here is not to enter these technical details, we will use the syntax provided in the vignette of the **rstpm2** package without questioning the choices made in terms of calculation of time periods.

The following lines of code create a new dataset `colon2` containing amongst others the variables `exitage` and `exityear` (respectively, age and calendar year at the end of follow-up), as well as `sex`, that will be used to get the appropriate mortality rate.

```
R> colon2 <- within(rstpm2::colon, {
+     status <- ifelse(surv_mm > 120.5, 1, status)
+     tm <- pmin(surv_mm, 120.5) / 12
+     exit <- dx + tm * 365.25
+     sex <- as.numeric(sex)
+     exitage <- pmin(floor(age + tm), 99)
+     exityear <- floor(yydx + tm)
+ })
```

Now, we can create the variable `rate` by merging `colon2` with `popmort`:

```
R> colon2 <- merge(colon2, popmort, by.x = c("sex", "exitage", "exityear"),
+     by.y = c("sex", "age", "year"))
R> head(colon2[, c("sex", "age", "stage", "status",
+     "exitage", "exityear", "tm", "rate")], 3)
```

```
  sex age    stage status exitage exityear        tm         rate
1   1  12  Distant      2      13     1991 1.6250000 0.0002300408
2   1  16 Regional      2      16     1987 0.7083333 0.0009003758
3   1  17  Distant      2      17     1979 0.1250000 0.0010305587
```

This dataset will be used later to compare `mexhaz()` to functions for excess hazard estimation from other packages.

Note that population lifetables for many countries can be accessed through the Human Mortality Database website (http://www.mortality.org/). In general, the mortality rate is expressed in number of events per person-year: it might be rescaled (e.g., number of events per person-month) to match the time scale chosen in a particular application but there is no need to convert it to a specific scale, as is for example the case when using the **relsurv** package (all time variables having to be expressed in days).

3.4. Mixed-effect excess hazard regression model

In order to fit a (possibly, excess) hazard regression model with a random effect, the argument `random` is used to specify the name of the covariate defining the cluster. By default in **mexhaz**, the number of nodes of adaptive Gauss-Hermite quadrature (AGHQ) is set to 10 but it can be modified through the argument `n.aghq`.

Here are the results obtained when fitting a mixed-effect excess hazard model with 10 quadrature nodes:

```
Call:
mexhaz(formula = Surv(timesurv, vstat) ~ agec + IsexH + nph(IsexH),
    data = simdatn1, expected = "popmrate", base = "exp.bs",
    degree = 3, knots = c(1, 5), random = "clust", n.aghq = 10)


Coefficients:
                  Estimate      StdErr  t.value   p.value
Intercept       -0.4742698   0.1087197  -4.3623 1.319e-05 ***
BS3.1           -1.3784399   0.1870137  -7.3708 2.051e-13 ***
BS3.2           -1.6597091   0.1838144  -9.0293 < 2.2e-16 ***
BS3.3           -1.5328453   0.3492171  -4.3894 1.166e-05 ***
BS3.4           -2.4781726   0.3718693  -6.6641 3.027e-11 ***
BS3.5           -2.4239934   0.4101604  -5.9099 3.710e-09 ***
agec             0.0470502   0.0018874  24.9280 < 2.2e-16 ***
IsexH            0.9820957   0.1218117   8.0624 9.801e-16 ***
IsexH*BS3.1      0.2137447   0.2271847   0.9408    0.3468
IsexH*BS3.2      0.1252103   0.2477386   0.5054    0.6133
IsexH*BS3.3     -0.7965804   0.5146900  -1.5477    0.1218
IsexH*BS3.4      0.6929658   0.5605545   1.2362    0.2165
IsexH*BS3.5     -0.0590359   0.6212609  -0.0950    0.9243
clust [log(sd)] -1.4182882   0.1363752 -10.3999 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Hazard ratios (for proportional effect variables):
        Coef     HR CI.lower CI.upper
agec  0.0471 1.0482   1.0443   1.0521
IsexH 0.9821 2.6700   2.1028   3.3903
```

```
log-likelihood: -5776.5972 (for 14 degree(s) of freedom)

number of observations: 4000, number of events: 3237
```

Using less quadrature points will decrease the time needed to compute the cluster-specific marginal likelihood but may result in a poor approximation of the total likelihood: this might in turn increase time to convergence and sometimes even hamper the convergence of the model.

```
          agec clust (sd)   log-lik      AIC time (sec) nlm code
1Quad-P  0.047       0.242 -5776.649 11581.30       91.90        1
5Quad-P  0.047       0.242 -5776.598 11581.19       54.11        1
10Quad-P 0.047       0.242 -5776.597 11581.19       56.68        1
```

When a random effect model is fitted, the object returned by `mexhaz()` includes an estimate of the logarithm of the standard deviation of the random effect (`clust [log(sd)]`) in the `coefficients` slot and the predicted cluster-specific shrinkage factors can be found in the `mu.hat` slot. These shrinkage factors are used by the `predict.mexhaz()` function to calculate cluster-specific hazard and survival values. If no cluster name is given, predictions are made for the value 0 of random effect (but it should be reminded that hazard and survival predictions at the mean value of the random effect are different from the marginal hazard and survival values obtained by integrating over the distribution of the random effect).

For example, the following lines of code show how to obtain the baseline excess hazard for men aged 70 in cluster 15:

```
R> PBsExcR.c15 <- predict(ModBsExc.10n, time.pts = MyTime,
+    data.val = MyData[2, ], cluster = "15")
```

And for men aged 70 with the value 0 for the random effect:

```
R> PBsExcR.0 <- predict(ModBsExc.10n, time.pts = MyTime,
+    data.val = MyData[2, ])
```

3.5. Convergence issues in practice

A common problem encountered by users of statistical package in the process of constructing a regression model is that of non convergence, i.e., the algorithm is not able to find the values of the parameters corresponding to the specified model. Although there might be structural reasons explaining why "the model does not converge" (such as non-identifiability or collinearity between covariates), we list here a few problems that can be encountered when using `mexhaz()` and that may be solved either by modifying some of the arguments of the function or by changing the parametrization of the model.

First of all, it is a fact that there is generally no algorithm that works for all optimization problems. The R statistical software includes various optimization algorithms, the most commonly used being `nlm()` and `optim()` (the latter allows the user to choose among several methods such as `Nelder-Mead` or BFGS). In order to take advantage of the availability of these different optimization procedure, the `mexhaz()` function allows the user to choose between these options through the arguments `fnoptim` (which can take values `"nlm"` and `"optim"`).

The `method` argument takes as possible values the names of the different methods available in `optim` (as there is only one method for `nlm`). Moreover, it is possible to add extra-arguments to `mexhaz()` to further customize the calls to `nlm` or `optim` (e.g., maximal number of iterations, gradient tolerance, etc.). Information on the convergence of the model is provided in the output of our function (slot `code`). It is worth noticing here that the value of `code` is the one returned by the optimization method chosen by the user. Consequently, convergence is indicated by `code = 1` when `nlm` (the default) is used and by `code = 0` when `optim` is used.

Among the frequent causes of non convergence is the choice of initial values. Indeed, the likelihood function might present local extrema or be almost flat in some regions of the parameter space so that depending on the initial values, the algorithm might find itself stuck in such an area. Sensitivity to initial values is more likely to happen when the complexity of the model increases: we showed for example that the mixed-effect excess mortality hazard models were sensitive to initial values (Charvat *et al.* 2016). In such cases, a common (and usually effective) practice is to fit a simple model first and use the estimated parameters (possibly rounded) as initial values for more complex models. Based on this practice, it appears that the convergence of mixed-effect excess hazard models can be greatly improved by first fitting a fixed-effect excess hazard model and then use the rounded estimated values as initial values (setting the initial value of the standard deviation of the random effect at a reasonably small positive value such as 0.5). Initial values are specified through the argument `init`.

Another problem that one might be faced with is related to the numerical approximation of the likelihood. If this approximation is too crude, the algorithm might fail to locate the maximum of the likelihood function (or this maximum might not exist for the approximated likelihood). In mixed-effect hazard models fitted with the `mexhaz()` function, two possible causes of inadequate approximation of the likelihood function are i) the approximation of the cumulative hazard for each individual by Gauss-Legendre quadrature (when *B*-splines of degree 2 or 3, or restricted cubic splines, are used to model the logarithm of the baseline hazard) and ii) the approximation of the cluster-specific marginal likelihood by adaptive Gauss-Hermite quadrature. The user can modify the number of Gauss-Legendre nodes (argument `n.gleg`) and of Gauss-Hermite nodes (argument `n.aghq`): increasing the number of nodes will allow better convergence of the models but will require more computational time (especially for the adaptive Gauss-Hermite quadrature).

In the specific context of excess hazard models, convergence problems might also be encountered if the total hazard observed in the study population is lower than the expected hazard (obtained through population mortality tables). Indeed, the excess hazard should become negative, which is not permitted by its parameterization (i.e., the excess hazard is constrained to be a positive function of time). In that very particular case, there is no other solution than to drop the expected hazard term and fit a model for the total hazard, as assuming an excess hazard on this population does not sound appropriate.

The last problem we will mention in this section is related to the scale of the variables used in the model, whether it be i) the time scale or ii) the scale of the covariates used in the linear predictor.

Concerning the time scale, it should be adapted to the event-generating mechanism: if only a few dozen cases happen each year, and time is expressed in days or months, the hazard rate will be very small. Consequently, a regression model based on such a time scale might fail to estimate correctly the hazard because optimization algorithms might not be able to find a

correct step size to reach the maximum likelihood value of the parameters. A simple solution to this problem is of course to rescale time so that the hazard is expressed on a meaningful scale: invariance of the likelihood function towards model reparametrization insures that we are estimating the same model. One should notice that such problems are specific to parametric and flexible hazard models: with the semi-parametric Cox model, only the order of the events is used in the estimation procedure so that the time scale is of no importance.

A similar problem occurs with covariates taking very large (age expressed in days, age squared, etc.) or very small values (e.g., concentration of biomarkers expressed in g/L). Remembering that parameter estimates corresponds to a one unit increase of the variable of interest, it is easy to imagine that the effect of age expressed in days is likely to be very small in most applications. Consequently, the associated parameter will be very small and, as before, the optimization algorithm might have problem to deal with parameters of widely different magnitude and find its way in the parameter space towards the maximum likelihood. Once again, a possible solution to this problem is to rescale the variables.

3.6. Model parameterization in mexhaz compared to other packages

We focus here on the two main extensions proposed in our package, namely the possibility to include time-dependent effects and a random effect, in comparison with two existing tools proposed in R by default, **mgcv** (Wood 2017) and **nlme** (Pinheiro, Bates, and R Core Team 2021). In our model parameterization (see Equation 10), a time-dependent effect for a given covariate is defined as an interaction between the functional form defining the baseline hazard and that covariate. The user can specify that time-dependent effects are to be fitted by using the special term `nph()` in the model formula. The `nph()` term takes as argument the names of the covariates (separated by a plus operator) for which a time-dependent effect is assumed; because this time-dependent effect uses that same function of time as the baseline hazard, there is no need for further model specification. This is one difference with, for example, the smoother functions `s()` in the package **mgcv** where the user can specify a different degree and a different number of knots for each covariate. One planned extension of our package is to allow for different functions of time for each time-dependent effect.

Regarding the inclusion of a random effect, we used a specific argument `random` which contains the name of the variable defining the cluster level within quotes, following the essence of **nlme**, even though the `mexhaz()` function does not require a formula syntax (because it currently allows only one random effect). It is planned to extend the `mexhaz()` function so that it allows for several random effects.

4. Comparison with other packages for survival analysis

In this section, we compare the results obtained with **mexhaz** and with other R packages for analysis of time-to-event data using real datasets provided with these packages. We start with a comparison of **mexhaz** with two other packages that can be used to fit flexible parametric hazard models (**flexsurv** and **rstpm2**), for both overall mortality and excess mortality. Then we compare **mexhaz** with two other packages that can be used to fit mixed-effect hazard regression models (**survival**, **frailtypack** and **parfm**)[1].

---

[1]The versions of the packages used to produce the results detailed in this section are as follows: **mexhaz** 1.10; **survival** 3.2-11; **rstpm2** 1.5.2; **frailtypack** 3.3.2; **flexsurv** 2.0; **parfm** 2.7.6, and **muhaz** 1.2.6.4.

4.1. Flexible parametric hazard models

In the following, we used the dataset `bc` from the **flexsurv** package: it contains survival times and vital status (`recyrs` and `censrec`, respectively) of 686 patients with primary node positive breast cancer, with a variable defining the prognosis group (`group`). We compared the results obtained with **mexhaz** to those obtained with **flexsurv** and **rstpm2** when fitting a flexible parametric hazard model for the overall mortality.

```
R> data("bc", package = "flexsurv")
R> head(bc, 3)


  censrec rectime group    recyrs
1       0    1342  Good 3.676712
2       0    1578  Good 4.323288
3       0    1760  Good 4.821918
```

For **flexsurv**, we fitted the model corresponding to the one with the lower Akaike Information Criterion (AIC) retained in (Jackson 2016). For the model fitted with **rstpm2**, we used five degrees of freedom (without counting the intercept) to model the log-cumulative hazard and as many coefficients to model the time-dependent effect of the variable `group`. For **mexhaz**, we fitted a model assuming a cubic B-spline for the logarithm of the baseline hazard with two knots located at the tertiles of the distribution of event times and a time-dependent effect of the variable `group`.

When predicting and plotting the survival (Figure 4) and the corresponding hazard (Figure 5) for each prognosis group, we observed good agreement between **mexhaz** and the other methods. For comparison, we also provided the non-parametric smooth hazard estimate obtained with the **muhaz** package (Hess and Gentleman 2021).

4.2. Flexible parametric excess hazard models

In this section we again compared **mexhaz** with **rstpm2** and **flexsurv**, but regarding their ability to estimate the excess hazard. We used the `colon` dataset from the package **rstpm2** which contains 15,564 observations on colon cancer patients. The `popmort` dataset from the same package provides the corresponding population mortality rates that can be used to fit an excess hazard model. We analyzed this dataset with the objective of estimating the time-dependent effect of stage at diagnosis (i.e., the stage-specific hazards are not constrained to be multiples of one another), and we compared the results obtained with the three packages by reporting the stage-specific net survival predicted from the excess hazard regression models.

As in the overall survival setting, we observed very similar patterns of stage-specific predicted net survival (Figure 6).

4.3. Flexible mixed-effect hazard regression models

In this section, we compare results obtained with four functions for flexible mixed-effect hazard modeling (namely, `frailtyPenal()` from **frailtypack**, `stpm2()` from **rstpm2**, as well as `parfm()` and `mexhaz()` from the equally-named packages) and with the `coxph()` function from the **survival** package. The latter function, when used with the `frailty()` term in the
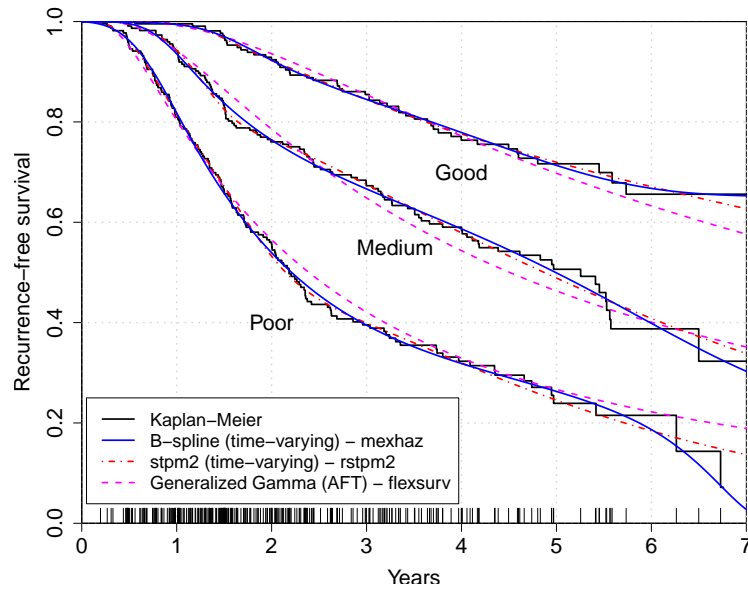
Figure 4: Comparison of the survival estimates provided by the different functions.
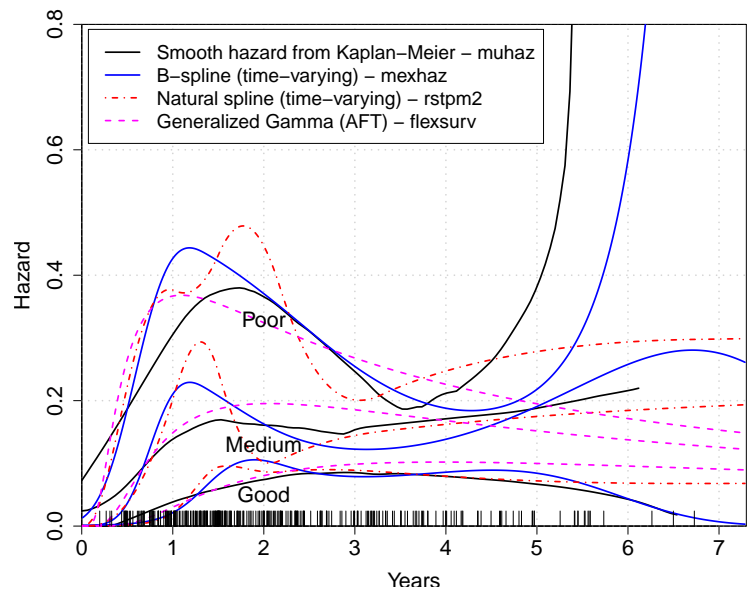

Figure 5: Comparison of the hazard estimates provided by the different functions.

model formula, allows the user to fit a shared frailty Cox model. The `simdatn2` dataset provided in the **mexhaz** package was used.

Concerning the hazard specification, we chose a Weibull parametric model for `parfm()`, and for the three packages allowing for flexible hazard modeling, we chose models with the same number (namely, seven) of degrees of freedom for the baseline hazard specification: for `mexhaz()`, the log-hazard was modeled with a cubic *B*-spline with three knots at the quartiles of the distribution of event times; for `frailtyPenal()`, five knots were specified for the
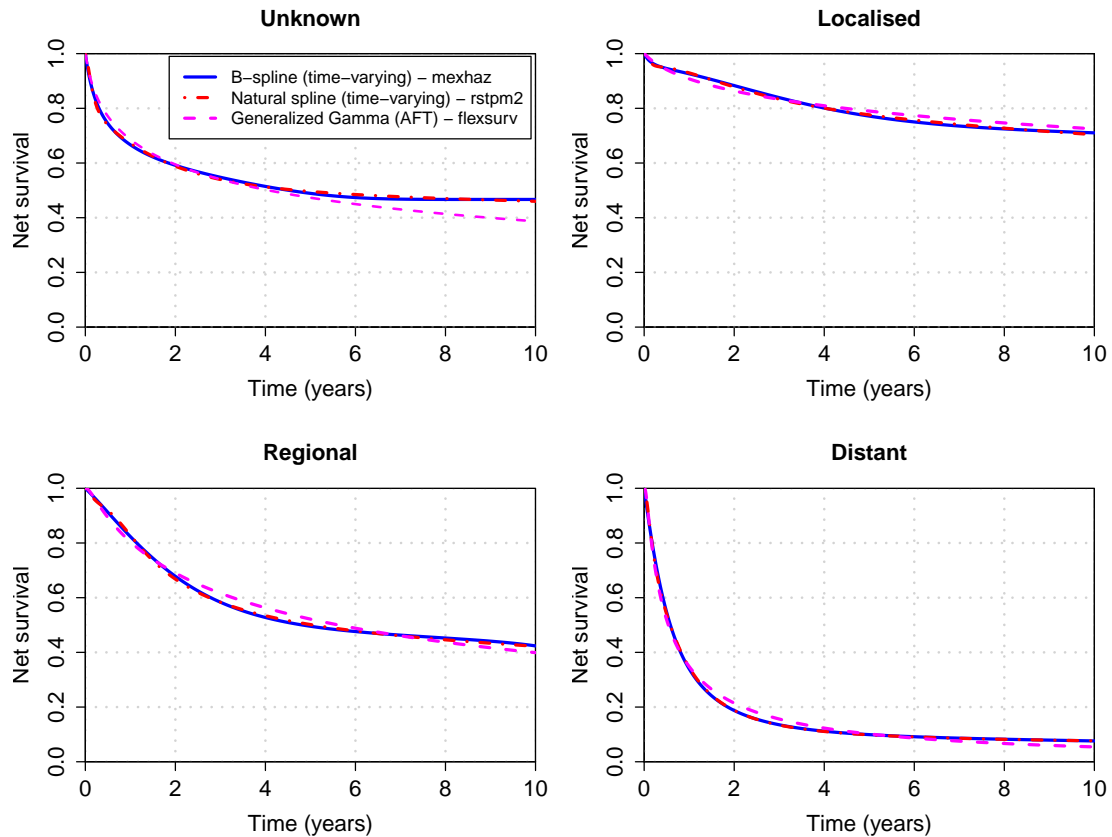
Figure 6: Comparison of net survival estimates given by the different functions.

cubic *M*-spline describing the hazard; and for `stpm2()`, the baseline cumulative hazard was modeled with a natural cubic spline with five knots.

The parameter estimates for the effect of `agecr` and `IsexH`, as well as their standard errors, were very similar between the different packages. We also observed a good agreement between the different packages regarding the estimated variance of the random-effect.

	agecr	sd(agecr)	IsexH	sd(IsexH)	sigma2	sd(sigma2)	time (sec)
coxph	5.3607	0.1699	0.9576	0.0376	0.0444	NA	0.1
mexhaz	5.3621	0.1702	0.9580	0.0377	0.0441	0.0120	7.7
stpm2	5.3656	0.1702	0.9580	0.0377	0.0441	0.0120	2.5
frailtyPenal	5.5498	0.1727	1.0377	0.0369	0.0697	0.0146	7.4
parfm	5.3095	0.1675	0.9489	0.0374	0.0430	0.0117	22.5

Finally, when comparing the shrinkage estimates (Figure 7), we also observed good concordance between **mexhaz**, **coxph**, **parfm** and **frailtypack**. However, for the latter, there seems to be a positivity constraint that forces negative shrinkage estimates to take on the value 0.
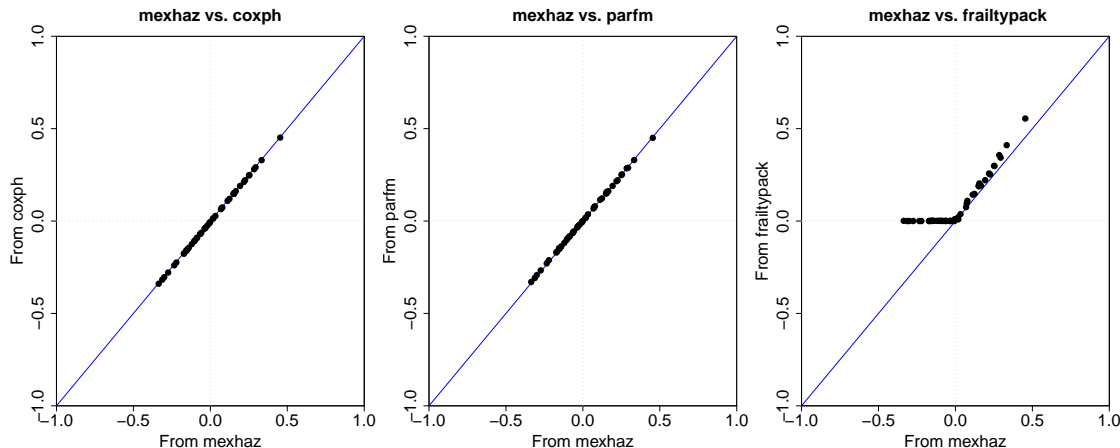
Figure 7: Comparison of values of the shrinkage estimates given by the different functions.

Note on execution time

As can be seen on the previous example, the execution time[2] varies between the functions (used here with their default internal specifications). From a general point of view, the execution time is related to various factors, some pertaining to the dataset (e.g., the number of individuals, the number of clusters), and others to the model specification (e.g., parametric model, flexible (log-)hazard model, flexible log-cumulative hazard, number of quadrature knots for approximating the cluster-specific marginal likelihood, etc.).

In order to give an example of how the execution time of the `mexhaz()` function scales up and compares with functions from other packages, we simulated a dataset of size 100,000 (500 balanced clusters of size 200 each) with the same methodology used to generate `simdatn1`. The comparison was restricted to `mexhaz()`, `stpm2()` and `frailtyPenal()` because they all can be used to fit flexible hazard models with a normally distributed random effect. Concerning the number of degrees of freedom to describe the baseline hazard, we used the same rules as described above. For the number of adaptive quadrature nodes, the default value of 20 nodes is used in `frailtyPenal()` for fitting spline-based hazard functions. Even if, as mentioned before, 10 nodes might be sufficient in most applications, we set the number of nodes in `mexhaz()` and `stpm2()` to 20 in order to allow for fair comparison. Results showed that `stpm2()` was the fastest (around 1 minute and 30 seconds), while `mexhaz()` was more than two times slower (3 minutes and 20 seconds), and `frailtyPenal()` nine times slower. The execution time observed with `stpm2()` can be explained by the fact that modeling is done on the log-cumulative hazard scale and thus does not require integration of the hazard as in `mexhaz()`, at the cost of the aforementioned limitations. On the other hand, the likelihood penalization process might explain the longer execution time observed with `frailtyPenal()`.

5. Discussion and conclusion

The **mexhaz** package combines different tools to model time-to-event data based on maximum likelihood theory, from flexible parametric models up to flexible parametric excess hazard

---

[2]Results were obtained using a computer with an Intel Core i7-9700 CPU with 16GB of RAM.

model including a random effect to analyze clustered data. Its implementation is efficient, computationally robust and compares very well with different existing packages devoted to some of the implemented features. The spline functions are powerful tools to provide smooth estimates for either the baseline hazard or the time-dependent effect of covariates, such effects being particularly important in areas such as cancer epidemiology. Even if the R code provided in the examples use a single time-dependent effect, extension to multiple time-dependent effects is straightforward. One feature of the package that we have not detailed here is the fact that the output of the `mexhaz()` and `predict.mexhaz()` functions can be used to derive other survival-related indicators. For example, Kipourou, Charvat, Rachet, and Belot (2019) show an example of use of **mexhaz** to derive the (adjusted) cumulative incidence functions with their confidence intervals in a competing risk setting.

We focused here on the free R software environment, but tools in Stata and SAS are also available for fitting (excess) hazard regression models with the possible inclusion of random effects. In Stata, the user-written commands **stpm2** (Lambert and Royston 2009) and **strcs** (Bower, Crowther, and Lambert 2016) can be used to fit flexible models defined on the cumulative hazard scale (the Royston-Parmar model) or the log-hazard scale, respectively, for both the overall mortality hazard and the excess mortality hazard. The commands **streg** and **mestreg** allow the user to fit parametric mixed-effect survival model with either a gamma or a log-normally distributed frailty, respectively (the command **mestreg** being however not restricted to two-level mixed effect models), while **stcox** allows the user to fit Cox proportional hazard models with a shared gamma frailty through the option `shared()`. The user-written **stmixed** (Crowther 2019) command provides a complementary program for fitting multilevel parametric survival models defined on the cumulative hazard scale, with the possibility to perform excess hazard modeling. In SAS, the procedure **PHREG** includes a `RANDOM` statement for specifying a shared frailty, the `DIST` option allowing this frailty to be gamma or log-normally distributed. Besides, the SAS procedure **NLMIXED** is an extremely powerful tool for developing two-level random effect models and can be used to construct shared frailty hazard-based regression model, where one can assume either a parametric distribution for the time-to-event (such as Weibull (Liu and Huang 2008; Kong, Archer, Moulton, Gray, and Wang 2010)) or different kinds of specification of the baseline hazard (such as piecewise constant (Dupont, Bossard, Remontet, and Belot 2013) or splines (Belot, Rondeau, Remontet, and Giorgi 2014)).

The current development of **mexhaz** allows the analysis of two-level hierarchical time-to-event data through the use of a random effect defined at the cluster level. Based on adaptive Gauss-Hermite quadrature, we showed that our approach provides reasonable estimates of the effects of the covariates defined either at the individual or cluster level, as long as enough clusters are present in the data (50 or more) (Charvat *et al.* 2016).

Acknowledgments

References

Balan T, Putter H (2019). "**frailtyEM**: An R Package for Estimating Semiparametric Shared Frailty Models." *Journal of Statistical Software*, **90**(7), 1–29. `doi:10.18637/jss.v090.i07`.

Belot A, Abrahamowicz M, Remontet L, Giorgi R (2010). "Flexible Modeling of Competing Risks in Survival Analysis." *Statistics in Medicine*, **29**(23), 2453–2468. `doi:10.1002/sim.4005`.

Belot A, Ndiaye A, Luque-Fernandez MA, Kipourou DK, Maringe C, Rubio FJ, Rachet B (2019). "Summarizing and Communicating on Survival Data According to the Audience: A Tutorial on Different Measures Illustrated with Population-Based Cancer Registry Data." *Clinical Epidemiology*, **11**, 53–65. `doi:10.2147/clep.s173523`.

Belot A, Rondeau V, Remontet L, Giorgi R (2014). "A Joint Frailty Model to Estimate the Recurrence Process and the Disease-Specific Mortality Process without Needing the Cause of Death." *Statistics in Medicine*, **33**(18), 3147–3166. `doi:10.1002/sim.6140`.

Booth JG, Hobert JP (1998). "Standard Errors of Prediction in Generalized Linear Mixed Models." *Journal of the American Statistical Association*, **93**(441), 262–272. `doi:10.1080/01621459.1998.10474107`.

Bossard N, Velten M, Remontet L, Belot A, Maarouf N, Bouvier AM, Guizard AV, Tretarre B, Launoy G, Colonna M, Danzon A, Molinie F, Troussard X, Bourdon-Raverdy N, Carli PM, Jaffré A, Bessaguet C, Sauleau E, Schvartz C, Arveux P, Maynadié M, Grosclaude P, Estève J, Faivre J (2007). "Survival of Cancer Patients in France: A Population-Based Study from The Association of the French Cancer Registries (FRANCIM)." *European Journal of Cancer*, **43**(1), 149–160. `doi:10.1016/j.ejca.2006.07.021`.

Bower H, Crowther MJ, Lambert PC (2016). "**strcs**: A Command for Fitting Flexible Parametric Survival Models on the Log-Hazard Scale." *Stata Journal*, **16**(4), 989–1012. `doi:10.1177/1536867x1601600410`.

Charvat H, Belot A (2021). **mexhaz**: *Mixed Effect Excess Hazard Models*. R package version 1.11, URL `https://CRAN.R-project.org/package=mexhaz`.

Charvat H, Remontet L, Bossard N, Roche L, Dejardin O, Rachet B, Launoy G, Belot A (2016). "A Multilevel Excess Hazard Model to Estimate Net Survival on Hierarchical Data Allowing for Non-Linear and Non-Proportional Effects of Covariates." *Statistics in Medicine*, **35**(18), 3066–3084. `doi:10.1002/sim.6881`.

Clements M, Liu XR, Christoffersen B (2021). **rstpm2**: *Smooth Survival Models, Including Generalized Survival Models*. R package version 1.5.2, URL `https://CRAN.R-project.org/package=rstpm2`.

Clerc-Urmès I, Grzebyk M (2020). **flexrsurv**: *Flexible Relative Survival Analysis*. R package version 1.4.5, URL `https://CRAN.R-project.org/package=flexrsurv`.

Cox DR, Oakes D (1984). *Analysis of Survival Data*. Chapman & Hall/CRC, Boca Raton.

Crowther MJ (2019). "Multilevel Mixed-Effects Parametric Survival Analysis: Estimation, Simulation, and Application." *Stata Journal*, **19**(4), 931–949. `doi:10.1177/1536867x19893639`.

Crowther MJ, Andersson TM, Lambert PC, Abrams KR, Humphreys K (2016). "Joint Modelling of Longitudinal and Survival Data: Incorporating Delayed Entry and an Assessment of Model Misspecification." *Statistics in Medicine*, **35**(7), 1193–1209. `doi:10.1002/sim.6779`.

Duchateau L, Janssen P (2008). *The Frailty Model*. Springer-Verlag.

Dupont C, Bossard N, Remontet L, Belot A (2013). "Description of an Approach Based on Maximum Likelihood to Adjust an Excess Hazard Model with a Random Effect." *Cancer Epidemiology*, **37**(4), 449–456. `doi:10.1016/j.canep.2013.04.001`.

Estève J, Benhamou E, Croasdale M, Raymond L (1990). "Relative Survival and the Estimation of Net Survival: Elements for Further Discussion." *Statistics in Medicine*, **9**, 529–538. `doi:10.1002/sim.4780090506`.

Geskus RB (2015). *Data Analysis with Competing Risks and Intermediate States*. Chapman & Hall/CRC.

Giner G, Smyth GK (2016). "**statmod**: Probability Calculations for the Inverse Gaussian Distribution." *R Journal*, **8**(1), 339–351. `doi:10.32614/rj-2016-024`.

Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Estève J, Gouvernet J, Faivre J (2003). "A Relative Survival Regression Model Using B-Spline Functions to Model Non-Proportional Hazards." *Statistics in Medicine*, **22**, 2767–2784. `doi:10.1002/sim.1484`.

Golub GH, Welsch JH (1969). "Calculation of Gauss Quadrature Rules." *Mathematics of Computation*, **23**(106), 221–230. `doi:10.1090/s0025-5718-69-99647-1`.

Haller B, Schmidt G, Ulm K (2013). "Applying Competing Risks Regression Models: An Overview." *Lifetime Data Analysis*, **19**(1), 33–58. `doi:10.1007/s10985-012-9230-8`.

Hess K, Gentleman R (2021). **muhaz**: *Hazard Function Estimation in Survival Analysis*. R package version 1.2.6.4, URL `https://CRAN.R-project.org/package=muhaz`.

Jackson C (2016). "**flexsurv**: A Platform for Parametric Survival Modeling in R." *Journal of Statistical Software*, **70**(1), 1–33. `doi:10.18637/jss.v070.i08`.

Kalbfleisch JD, Prentice RL (2002). *The Statistical Analysis of Failure Time Data*. 2nd edition. John Wiley & Sons, New York. `doi:10.1002/9781118032985`.

Kipourou DK, Charvat H, Rachet B, Belot A (2019). "Estimation of the Adjusted Cause-Specific Cumulative Probability Using Flexible Regression Models for the Cause-Specific Hazards." *Statistics in Medicine*, **38**(20), 3896–3910. `doi:10.1002/sim.8209`.

Kong X, Archer KJ, Moulton LH, Gray RH, Wang MC (2010). "Parametric Frailty Models for Clustered Data with Arbitrary Censoring: Application to Effect of Male Circumcision on HPV Clearance." *BMC Medical Research Methodology*, **10**(1), 40. `doi:10.1186/1471-2288-10-40`.

Lambert PC, Royston P (2009). "Further Development of Flexible Parametric Models for Survival Analysis." *Stata Journal*, **9**(2), 265–290. doi:10.1177/1536867x0900900206.

Liu L, Huang X (2008). "The Use of Gaussian Quadrature for Estimation in Frailty Proportional Hazards Models." *Statistics in Medicine*, **27**(14), 2665–2683. doi:10.1002/sim.3077.

Liu Q, Pierce DA (1994). "A Note on Gauss-Hermite Quadrature." *Biometrika*, **81**, 624–629. doi:10.2307/2337136.

Mahboubi A, Abrahamowicz M, Giorgi R, Binquet C, Bonithon-Kopp C, Quantin C (2011). "Flexible Modeling of the Effects of Continuous Prognostic Factors in Relative Survival." *Statistics in Medicine*, **30**(12), 1351–1365. doi:10.1002/sim.4208.

Mathews JH, Fink KD (1999). *Numerical Methods Using* MATLAB. 3rd edition. Simon & Schuster.

Munda M, Rotolo F, Legrand C (2012). "**parfm**: Parametric Frailty Models in R." *Journal of Statistical Software*, **51**(1), 1–20. doi:10.18637/jss.v051.i11.

Nelson CP, Lambert PC, Squire IB, Jones DR (2007). "Flexible Parametric Models for Relative Survival, with Application in Coronary Heart Disease." *Statistics in Medicine*, **26**, 5486–5498. doi:10.1002/sim.3064.

Pinheiro J, Bates D, R Core Team (2021). **nlme**: *Linear and Nonlinear Mixed Effects Models*. R package version 3.1-152, URL https://CRAN.R-project.org/package=nlme.

Pinheiro JC, Bates DM (1995). "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model." *Journal of Computational and Graphical Statistics*, **4**(1), 12–35. doi:10.2307/1390625.

Pohar Perme M, Henderson R, Stare J (2009). "An Approach to Estimation in Relative Survival Regression." *Biostatistics*, **10**, 136–146. doi:10.1093/biostatistics/kxn021.

Pohar Perme M, Pavlič K (2018). "Nonparametric Relative Survival Analysis with the R Package **relsurv**." *Journal of Statistical Software*, **87**(8), 1–27. doi:10.18637/jss.v087.i08.

Pohar Perme M, Stare J (2006). "Relative Survival Analysis in R." *Computer Methods and Programs in Biomedicine*, **81**, 272–278. doi:10.1016/j.cmpb.2006.01.004.

Pohar Perme M, Stare J (2007). "Making Relative Survival Analysis Relatively Easy." *Computers in Biology and Medicine*, **37**, 1741–1749. doi:10.1016/j.compbiomed.2007.04.010.

Pohar Perme M, Stare J, Estève J (2012). "On Estimation in Relative Survival." *Biometrics*, **68**(1), 113–120. doi:10.1111/j.1541-0420.2011.01640.x.

Putter H, Fiocco M, Geskus R (2007). "Tutorial in Biostatistics: Competing Risks and Multi-State Models." *Statistics in Medicine*, **26**, 2389–2430. doi:10.1002/sim.2712.

Quantin C, Abrahamowicz M, Moreau T, Bartlett G, MacKenzie T, Adnane Tazi M, Lalonde L, Faivre J (1999). "Variation Over Time of the Effects of Prognostic Factors in a Population-Based Study of Colon Cancer: Comparison of Statistical Models." *American Journal of Epidemiology*, **150**(11), 1188–1200. doi:10.1093/oxfordjournals.aje.a009945.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid N (1994). "A Conversation with Sir David Cox." *Statistical Science*, **9**(3), 439–455. doi:10.1214/ss/1177010394.

Remontet L, Bossard N, Belot A, Estève J (2007). "An Overall Strategy Based on Regression Models to Estimate Relative Survival and Model the Effects of Prognostic Factors in Cancer Survival Studies." *Statistics in Medicine*, **26**(10), 2214–2228. doi:10.1002/sim.2656.

Rondeau V, Marzroui Y, Gonzalez J (2012). "**frailtypack**: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation." *Journal of Statistical Software*, **47**(1), 1–28. doi:10.18637/jss.v047.i04.

Royston P, Lambert PC (2011). *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. StataCorp LP.

Royston P, Parmar MKB (2002). "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. doi:10.1002/sim.1203.

Therneau TM (2020). **coxme**: *Mixed Effects Cox Models*. R package version 2.2-16, URL https://CRAN.R-project.org/package=coxme.

Therneau TM (2021). **survival**: *Survival Analysis*. R package version 3.2-11, URL https://CRAN.R-project.org/package=survival.

Van den Berg G, Drepper B (2016). "Inference for Shared-Frailty Survival Models with Left-Truncated Data." *Econometric Reviews*, **35**(6), 1075–1098. doi:10.1080/07474938.2014.975640.

Wienke A (2010). *Frailty Models in Survival Analysis*. Chapman & Hall/CRC.

Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2nd edition. Chapman & Hall/CRC. doi:10.1201/9781315370279.

Affiliation:

Hadrien Charvat
Division of International Collaborative Research
Center for Public Health Sciences
National Cancer Center
Tokyo, Japan
E-mail: chadrien@ncc.go.jp

Aurélien Belot
Inequalities in Cancer Outcomes Network
Faculty of Epidemiology and Population Health
Department of Non-Communicable Disease Epidemiology
London School of Hygiene & Tropical Medicine
London, United Kingdom
E-mail: Aurelien.Belot@lshtm.ac.uk