Andreas Bender

Flexible Modeling of Time-to-event Data and Exposure-Lag-Response Associations

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 05. Juni 2018

Andreas Bender

Flexible Modeling of Time-to-event Data and Exposure-Lag-Response Associations

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 05. Juni 2018

Erste Berichterstatterin: Prof. Dr. Helmut Küchenhoff
Zweiter Berichterstatter: Prof. Dr. Jan Beyersmann
Dritter Berichterstatter: Prof. Dr. Andreas Groll

Tag der Disputation: 27.07.2018

Acknowledgments

I thank Prof. Dr. Helmut Küchenhoff for giving me the opportunity to pursue my PhD under his supervision, and his support and encouragement throughout the years. Working at the StaBLab provided the perfect mix of research, teaching and applied statistics.

I wish I could include a GIF to show my gratitude towards Dr. Fabian Scheipl, who supported me in almost all of my research projects and was the perfect collaborator.

I thank Prof. Dr. Wolfgang Hartl for initiating the research project that led to this thesis and for his collaboration on multiple successful projects over the past years.

I thank Prof. Dr. Jan Beyersmann for serving as a second reviewer and hosting me as a speaker at one of the conferences at the University of Ulm early into my PhD.

Many thanks to Prof. Dr. Andreas Groll for serving as a reviewer and for encouraging me to pursue a PhD after I obtained my Masters degree.

I would also like to thank

... Prof. Dr. Göran Kauermann for serving as chair at my disputation and Prof. Dr. Thomas Augustin for being available as a backup reviewer.

... Prof. Dr. Anne-Laure Boulesteix who gave me the opportunity to work on interesting research projects back when I was an undergraduate student and thus was ultimately responsible for my way into academia.

... all my colleagues at the StaBLab, especially Andre for the many ours teaching **R** courses, Alex for collaborating on the KOALA project and Felix for the work on propensity scores.

... my friends and colleagues David, Schorschi, Norbert, Verena and Gunther, for many fun ours at the gym, at restaurants, at the local bars, on mountain tops and on bike saddles.

Finally and most importantly, I would like to thank my parents Frieda and Alexander, for their unmeasurable and unwavering support throughout the years.

Summary

This thesis introduces a general framework for the representation and estimation of so-called exposure-lag-response associations, in the context of time-to-event data analysis. The development of this framework was motivated by the research question how calorie intake is associated with survival of critically ill patients. A second focus of the work was on making these methods accessible to applied scientists.

In the first part of this work the use of generalized additive mixed models for time-to-event data analysis (without cumulative effects) is introduced, focusing on accessibility, practical applications and smooth, time-varying effects.

The second part defines the general framework for the estimation of complex exposure-lag-response associations. While Chapter 3 focuses on the methodological development of the methods and provides extensive simulation studies, Chapter 4 reevaluates the analysis addressed to a medical audience and sets a stronger focus on the application and practically relevant sensitivity and subgroup analyses.

In the third part, the open source **R**-package **pammtools** is introduced. **pammtools** implements functions that facilitate the estimation and visualization of the methods discussed in the first two parts. The functionality is described in detail using real and simulated data. Whenever applicable, the discussed methods are compared to other established estimation frameworks for time-to-event data analysis.

Zusammenfassung

In dieser Arbeit wird eine allgemeiner Rahmen zur Darstellung und Schätzung sogenannter *exposure-lag-response* Assoziazionen im Kontext von Überlebensdaueranalysen vorgestellt. Die Entwicklung dieser Methoden wurde motiviert durch die Forschungsfrage wie Kalorienzufuhr mit dem Überleben von Intensivpatienten assoziiert ist. Ein weiterer Fokus dieser Arbeit lag darin diese Methoden allgemein zugänglich zu machen.

Im ersten Teil der Arbeit wird die Verwendung von generalisierten additiven gemischten Modellen für Überlebensdaueranalysen (ohne kumulative Effekte) vorgestellt. Der Fokus der Darstellung liegt dabei auf Verständlichkeit, praktischen Anwendungen und glatten, zeit-variierenden Effekten.

Im zweiten Teil wird die allgemeine Darstellung zur Schätzung von komplexen kumulativen Effekten definiert. Während der Fokus von Kapitel 3 auf der methodischen Entwicklung, unterstützt durch extensive Simulationsstudien, liegt, werden in Kapitel 4 die Analysen auf klinische Anwender ausgerichtet und mit einem stärkeren Fokus auf die Anwendung und praxisrelevante Sensitivitäts- und Subgruppenanalysen reevaluiert.

Im dritten Teil wird das Open Source **R** Pakett **pammtools** vorgestellt. **pammtools** implementiert Fuktionen, die die Schätzung und Visualisierung der in den ersten zwei Teilen dargestellten Methoden unterstützt. Die Funktionalität des Pakets wird anhand echter und simulierter Daten detailiert beschrieben. Wann immer möglich, werden die vorgestellten Methoden mit anderen etablierten Schätzmethoden für Überlebendauern verglichen.

Contents

Chapter 1

Introduction

1.1 Overview

This chapter summarizes the work presented in this thesis and gives an overview of the different contributing publications. Doing so, some more emphasis is put on the historical development of the methods presented in this work and some topics are discussed in more detail. Section 1.2 introduces the general setup and notation for the following sections. In Section 1.3 Generalized Additive Mixed Models (GAMMs) are introduced and some details on inference procedures relevant for Piece-wise exponential additive mixed models (PAMMs) are discussed in more detail. Section 1.4 reviews the motivation for the Piece-wise Exponential Model (PEM). In Section 1.5 the transition from PEMs to PAMMs is briefly recapped. Section 1.6 introduces cumulative effects and gives a review of the literature on this topic. In Section 1.7 the calculation of standard errors and confidence intervals for quantities derived from the log-hazard estimated by PAMMs is discussed in more detail. Finally, in Section 1.8, simulation studies are performed to investigate the sensitivity of PAMMs to the split point selection and to compare the different methods for uncertainty quantification discussed in Section 1.7.

The methodological development in this work was initiated by a cooperation between the statistical consulting unit (StaBLab) and Klinikum Großhadern. It was motivated by the clinically important question how nutrition, or more specifically calorie intake, affects the survival of patients during their stay in the intensive care unit (ICU) and hospital. Even though many studies on this topic have been published in the relevant literature, their interpretation and application to clinical practice remains controversial. For example, Patel et al. (2017) just recently discussed the key difference between the Canadian and American guidelines with respect to nutrition in clinical practice. One reason why study results and their applicability are discussed controversially, is that in many studies only patients from highly selected subgroups, e.g., only cancer patients, are included in the analysis, which precludes conclusions regarding other patient subgroups. Another reason is the high variability in analytical strategies including the definition of end-points, inclusion and modeling of confounders and overall analytic strategy. This is pointedly summarized by the publication “Optimal amount of

calories for critically ill patients: Depends on how you slice the cake!"(Heyland et al., 2011). The choice of analytical strategy is especially sensitive in the context of observational studies, where, among many other possible complications, complex adjustment for confounding is usually paramount. On the upside, observational studies have the advantage that a more heterogeneous population is available for analysis, potentially observed under more realistic circumstances compared to clinical trials.

One particular point that is often neglected when such data is analyzed is the dynamic nature of an ICU stay. For example, the calorie intake varies throughout the patient's stay in the ICU. Furthermore, it is reasonable to assume that not only the current calorie intake affects a patient's outcome, but rather that all (or some) past intakes affect the outcome cumulatively. One particular focus of this work was thus to develop methods that can flexibly estimate such effects in the context of time-to-event data analysis and to provide a general framework that can be applied to other studies in different fields of research. Another focus was on making these methods available to other researchers and applied scientists by providing an open source implementation of these methods. This research was facilitated by a large international data base (cf. Alberda et al., 2009) of almost 13000 ICU patients from 451 ICUs collected over the years 2007, 2008, 2009 and 2012. The pre-processed data used for the analysis in Bender et al. (2018b) is openly available at `https://doi.org/10.5281/zenodo.1203238`.

At this point it is important to note, that some terminology used in the context of survival analysis has a different meaning in the context of generalized regression models. Most notably, the Cox regression model (1.14) is often referred to as semi-parametric, because it is comprised of a parametric part (covariate effects) and a non-parametric part (baseline hazard). Also, additive models mostly refer to models where the covariate effects contribute to the hazard additively, e.g., the Aalen model (Aalen, 1978; Martinussen and Scheike, 2006). In the context of generalized regression models on the other hand, covariate effects are referred to as semi-parametric when the effect of one covariate is represented by a linear combination of multiple coefficients, e.g., (1.5), and models for the penalized estimation of such effects are referred to as additive models. In this work, and in the contributing articles, the latter convention is used, if not explicitly specified otherwise.

1.2 Prerequisites

In this work we consider right censored time-to-event data. The random variable of continuous event times is denoted by T and censoring times are denoted by C. Observation units are denoted by $i = 1, \ldots, n$. The i-th observed event time is defined by $t_i = \min(T_i, C_i)$, the minimum of the event time T_i and censoring time C_i. Whether the event time of subject i was observed fully or censored is denoted by $\delta_i = I(T_i \leq C_i)$. In the following, we assume that $T_i \perp C_i, \forall i = 1, \ldots, n$ (conditional on covariates).

Let $t \sim F(t)$ the cumulative distribution function with density function $f(t)$. The *survival probability* (also survivor function) is given by

$$S(t) = 1 - F(t) = \mathbb{P}(T > t) \tag{1.1}$$

and the *hazard rate* is defined as

$$\lambda(t) = \lim_{\Delta_t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta_t) | T \geq t)}{\Delta_t} = \frac{f(t)}{S(t)} = -\frac{\mathrm{d} \log(S(t))}{\mathrm{d}t}. \tag{1.2}$$

Another quantity often of interest in time-to-event data analysis is the *cumulative hazard rate*, which can be defined in terms of the hazard rate (1.2) or the survival probability (1.1)

$$\Lambda(t) = \int_0^t \lambda(s) \mathrm{d}s = -\log(S(t)). \tag{1.3}$$

1.3 Generalized Additive Mixed Models

As will be shown later (Section 1.4), PEMs, and therefore PAMMs, can be specified as a generalized additive mixed Poisson regression models. Thus the estimation and inference procedures developed for GAMMs can mostly be applied directly to PAMMs. In this section, the methodological framework of GAMMs is briefly summarized, mostly following the exposition in (Wood, 2017, Ch. 6).

The general model formulation is given in equation (1.4)

$$\xi(\mu_i) = \eta_i = \tilde{\mathbf{X}}_i \boldsymbol{\beta} + \sum_{p=1}^{P} L_{i,p} f_p(x_p), y_i \sim F(\mu_i, \phi), \tag{1.4}$$

where ξ is the link function that maps the expectation μ_i to the linear predictor η_i. In general, F can denote the cumulative distribution function of any exponential family distribution, however, in this work we consider Poisson distributed (pseudo-observations) $y_i \sim Po(\mu_i)$. For the Poisson distribution $\phi = 1$ and thus will be omitted from the notation in the following. $\tilde{\mathbf{X}}_i \boldsymbol{\beta}$ summarizes all strictly parametric components of the model, while $\sum_p L_{i,p} f_p(x_p)$ is the sum of all linear functionals. Here, $p = 1, \ldots, P$ is an index for the p-th smooth, x_p denotes the possibly vector valued covariates associated with the p-th smooth and $L_{i,p}$ is a linear functional that maps x_p to scalar values. In the simplest case, $L_{i,p}$ could be a simple evaluation functional, such that $L_{i,p} f_p(x_p) = f_p(x_{i,p})$. Special cases of the linear functional most relevant in this work are cumulative effects described in Section 1.5 (integration functional).

Each smooth function is represented by a suitable basis function expansion, e.g., equations (1.5) and (1.6) show the basis function expansions for a univariate and a bivariate smooth respectively:

$$f_p(x) = \sum_{k=1}^{K_p} \gamma_{p,k} B_{p,k}(x), \tag{1.5}$$

$$f_p(x_1, x_2) = \sum_{m=1}^{M_p} \sum_{k=1}^{K_p} \gamma_{p,m,k} B_{p,m}(x_1) B_{p,k}(x_2). \tag{1.6}$$

In (1.5) and (1.6) $B_{p,k}$ and $B_{p,m}$ are suitable basis functions (see Wood (2017, Ch. 5) for an overview of different options, e.g., B-Splines (Eilers and Marx, 1996)), and $\gamma_{p,k}$, $\gamma_{p,k,m}$ the respective basis coefficients, estimated from the data. The (marginal) basis dimensions of the p-th smooth are given by K_p and M_p, respectively.

Equation 1.6 represents an anisotropic tensor product smooth (Wood, 2017, Ch. 5.6), which can be thought of as a non-linear, smooth interaction of two continuous covariates. In the context of PAMMs these can be used to represent non-linear, non-linearly time-varying effects $f_p(t, x)$. Anisotropic means that different degrees of smoothness are permitted for each dimension, which is useful, when the two covariates constituting the interaction term are measured on different scales, in different units, or might have different degrees of smoothness. Such tensor product smooths will be used in Section 1.6 to define the three-variate cumulative effects of time-dependent covariates.
By representing smooth functions through basis expansions, the resulting model is linear in the parameters and as such can be estimated as a standard generalized linear model (GLM), e.g., by Fisher Scoring. However, the choice of basis dimensions K_p and M_p makes the estimation somewhat arbitrary and the models are prone to over-fitting, especially for large K_p and M_p, respectively, therefore a quadratic penalty of the basis coefficients is included for each smooth, as described in the following.

Let $\mathbf{B}_p$ the design matrix associated with the p-th smooth, with basis functions $B_{p,k}(x_p), k = 1, \ldots, K_p$ evaluated at the observed values of $\mathbf{x}_p$ and $\boldsymbol{\gamma}_p = (\gamma_1, \ldots, \gamma_{K_p})'$ the vector of basis coefficients of the p-th smooth, such that $f_p(x_p) = \sum_k \gamma_{p,k} B_{p,k}(x_p) = \mathbf{B}_p \boldsymbol{\gamma}_p$. The coefficients of each smooth are penalized by a quadratic penalty of the form $\boldsymbol{\gamma}_p' \tilde{\mathbf{S}}_p \boldsymbol{\gamma}_p$, where $\tilde{\mathbf{S}}_p$ is suitably chosen depending on the type of smooth. The complete model matrix for model (1.4) can thus be obtained by column-wise combination of the individual design matrices $\mathbf{X} = (\tilde{\mathbf{X}}\ \mathbf{B}_1\ \cdots\ \mathbf{B}_P)$ with associated coefficient vector $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_P')'$. The total penalty over all smooth terms is given by $\mathbf{S}_\nu = \sum_{p=1}^{P} \nu_p \boldsymbol{\gamma}' \mathbf{S}_p \boldsymbol{\gamma}$. Note that here $\mathbf{S}_p$ is $\tilde{\mathbf{S}}_p$ padded with zeros such that all contributions of coefficients not associated with the p-th smooth are set to zero (cf. Wood (2017, p. 176) for a motivation). Parameters ν_p are the smoothing parameters of the p-th smooth and control the smoothness of the term, i.e., $f_p \to 0$ for $\nu_p \to \infty$ and f_p unpenalized when $\nu_p = 0$.

With above notation, the optimization criterion for model $\xi(\mu_i) = \mathbf{X}_i\boldsymbol{\gamma},\ y_i \sim Po(\mu_i)$ is given by the penalized log-likelihood

$$l(\boldsymbol{\gamma}) - \frac{1}{2}\sum_{p=1}^{P} \nu_p \boldsymbol{\gamma}' \mathbf{S}_p \boldsymbol{\gamma}. \tag{1.7}$$

There are two parameter vectors that must be optimized in (1.7), (a) the coefficient vector $\boldsymbol{\gamma}$ and (b) the smoothing parameters $\boldsymbol{\nu} = \nu_1, \ldots, \nu_P$. When smoothing parameters $\nu_1, \ldots, \nu_P$ are fixed, the estimation of $\boldsymbol{\gamma}$ can be performed using penalized, iteratively re-weighted least squares (P-IRLS; Wood, 2017, Ch. 6.1.1). Let ξ the link function, such that $\xi(\mu_i) = \eta_i$ and $V(\mu_i)$ the exponential family specific variance function. Under assumption of the Poisson distribution $V(\mu_i) = \mu_i$. Let further $\alpha(\mu_i) = (1 + (y_i - \mu_i)(V'(\mu_i)/V(\mu_i) + \xi''(\mu_i)/\xi'(\mu_i)))$. The P-IRLS estimate for $\hat{\boldsymbol{\gamma}}$ can be obtained by algorithm 1.

1. Initialize expectation $\hat{\mu}_i = y_i$ and linear predictor $\hat{\eta}_i = \xi(\hat{\mu}_i)$
2. Iterate until convergence
   (a) Calculate $\tilde{y}_i = \xi'(\hat{\mu}_i)(y_i - \hat{\mu}_i)/\alpha(\hat{\mu}_i) + \hat{\eta}_i$ and $w_i = \alpha(\hat{\mu}_i)/(\xi'(\hat{\mu}_i)^2 V(\hat{\mu}_i))$
   (b) Obtain $\hat{\boldsymbol{\gamma}}$ from $\underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{ (\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\gamma})' \mathbf{W} (\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\gamma}) + \sum_{p=1}^{P} \nu_p \boldsymbol{\gamma}' \mathbf{S}_p \boldsymbol{\gamma} \right\}$, where $\tilde{\mathbf{y}} = \tilde{y}_i, \ldots, \tilde{y}_n$ is the vector of pseudo observations $\tilde{y}_i$ calculated in the previous step and $\mathbf{W}$ is a diagonal $n \times n$ matrix with elements $\mathbf{W}_{i,i} = w_i$.
   (c) Update $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\gamma}}$

Algorithm 1: P-IRLS algorithm for the estimation of $\hat{\boldsymbol{\gamma}}$.

For the estimation of $\boldsymbol{\nu}$, Wood (2017, Ch. 6.2) discusses different criteria, among others the UBRE score (Wood, 2017, Ch. 6.2.1), (generalized) cross validation (Wood, 2017, Ch. 6.2.2 and 6.2.3) as well as marginal (ML) and restricted maximum likelihood (REML) approaches (Wood, 2017, Ch. 6.26, 6.52), which is the suggested approach in Wood (2011).

The REML criterion for the optimization of $\boldsymbol{\nu}$ can be derived by taking the Mixed Model perspective of GAMMs that results from viewing the regression coefficients $\boldsymbol{\gamma}$ as random effects with zero mean and variance $\mathbf{S}_\nu^-$ drawn from an improper, multivariate normal distribution $\boldsymbol{\gamma} \sim N(\mathbf{0}, \mathbf{S}_\nu^-)$ with density (1.8), where $\mathbf{S}_\nu^-$ is a generalized inverse of $\mathbf{S}_\nu$.

$$f(\boldsymbol{\gamma}) = \frac{|\mathbf{S}_\nu|_+^{1/2}}{\sqrt{(2\pi)^{dim(\boldsymbol{\gamma})-M}}} \exp\left(\frac{-\boldsymbol{\gamma}' \mathbf{S}_\nu \boldsymbol{\gamma}}{2}\right) \tag{1.8}$$

In (1.8) $|\mathbf{S}_\nu|_+$ is a generalized determinant and M is the dimension of the null space of $\mathbf{S}_\nu$.

The restricted Likelihood is then obtained by integrating $\boldsymbol{\gamma}$ out of the $f(\mathbf{y}, \boldsymbol{\gamma})$, as shown below:

$$\int f(\mathbf{y}, \boldsymbol{\gamma}) \mathrm{d}\boldsymbol{\gamma} \simeq \int \exp\left[\log(f(\mathbf{y}, \hat{\boldsymbol{\gamma}})) + \frac{1}{2}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})' \frac{\partial^2 \log(f(\mathbf{y}, \boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})\right] \mathrm{d}\boldsymbol{\gamma} \tag{1.9}$$

$$= f(\mathbf{y}|\hat{\boldsymbol{\gamma}}) f(\hat{\boldsymbol{\gamma}}) \int \exp\left[\frac{1}{2}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})' \frac{\partial^2 \log(f(\mathbf{y}|\boldsymbol{\gamma})) + \log(f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})\right] \mathrm{d}\boldsymbol{\gamma} \tag{1.10}$$

$$= L(\hat{\boldsymbol{\gamma}}) f_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}) \int \exp\left[-\frac{1}{2}(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})'(\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{S}_\nu)(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})\right] \mathrm{d}\boldsymbol{\gamma} \tag{1.11}$$

$$= L(\hat{\boldsymbol{\gamma}}) f_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}) \frac{\sqrt{(2\pi)^{dim(\boldsymbol{\gamma})}}}{|\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{S}_\nu|^{1/2}}. \tag{1.12}$$

The second term in (1.9) is obtained by Taylor expansion around $\hat{\boldsymbol{\gamma}}$. In (1.10) the standard result $f(y, \boldsymbol{\gamma}) = f(y|\boldsymbol{\gamma}) f(\boldsymbol{\gamma})$ is used twice. For (1.11) the second derivative of the logarithmic density (1.8) is straight forward, while $\frac{\partial^2 l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}' \partial \boldsymbol{\gamma}} = -\mathbf{X}'\mathbf{W}\mathbf{X}$ is the negative, weighted Hessian matrix and $\mathbf{W}$ is the diagonal matrix of newton weights (from the P-IRLS step conditional on the current values of $\boldsymbol{\nu}$). The integral in (1.12) is obtained by integrating the kernel of a multivariate normal (cf. Wood (2017, Ch. 2.4) for details).

Finally, the REML criterion (1.13) suggested by Wood (2011) for the estimation of $\boldsymbol{\nu}$ follows by applying the logarithm to (1.12):

$$\mathcal{V}(\boldsymbol{\nu}) \simeq l(\hat{\boldsymbol{\gamma}}) - \frac{\hat{\boldsymbol{\gamma}}'\mathbf{S}_\nu\hat{\boldsymbol{\gamma}}}{2} - \frac{\log|\mathbf{S}_\nu|_+}{2} - \frac{\log|\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{S}_\nu|}{2} + \frac{M}{2}\log 2\pi, \tag{1.13}$$

where, $\hat{\boldsymbol{\gamma}}$ is the optimizer of the penalized log-likelihood (1.7) (or respective starting values).
To guarantee convergence, Wood (2011) suggests to run an outer iteration to estimate the smoothing parameters by optimizing the REML criterion (1.13) followed by an inner P-IRLS iteration (Wood, 2017, Ch. 6.1.1) to obtain $\hat{\boldsymbol{\gamma}}$ conditional on the current value of $\boldsymbol{\nu}$. The two steps are repeated until a pre-specified convergence criterion is met.

1.4 Piece-wise Exponential Model

Derivations of the PEM usually start with the general proportional hazards model (1.14)

$$\lambda(t; \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}_i'\boldsymbol{\beta}), \tag{1.14}$$

where the hazard at time t depends on $\mathbf{x}_i'$ the $1 \times P$ row-vector of covariates of subject i. Cox (1972) proposed to estimate parameters $\boldsymbol{\beta}$ of this model by optimizing the partial likelihood, while the baseline hazard $\lambda_0(t)$ can be left unspecified and estimated nonparametrically, e.g., by the Nelson-Aalen estimator or variants thereof.

A special case of (1.14) is obtained when the follow-up is partitioned in J intervals with interval boarders $\kappa_0 < \cdots < \kappa_J$, such that the j-th interval is defined by $(\kappa_{j-1}, \kappa_j]$, and the baseline hazard is assumed to be constant within each interval, i.e., $\lambda_0(t) = \lambda_j \ \forall t \in (\kappa_{j-1}, \kappa_j]$. In this case,

$$\lambda(t; \mathbf{x}_i) = \lambda_j \exp(\mathbf{x}_i' \boldsymbol{\beta}) =: \lambda_{ij} \ \forall t \in (\kappa_{j-1}, \kappa_j] \tag{1.15}$$

Let $j(i)$ be the index of the interval for which $t_i \in (\kappa_{j(i)-1}, \kappa_{j(i)}]$ and $\delta_{ij} \in \{0, 1\}$ the event indicator of subject i in interval j with $\delta_{ij(i)} = \delta_i$. Assuming (1.15), the log-likelihood contribution ℓ_i of observation unit i is given by

$$\begin{aligned}
\ell_i(\boldsymbol{\beta}) &= \log(\lambda(t_i; \mathbf{x}_i)^{\delta_i} S(t_i; \mathbf{x}_i)) && (1.16)\\
&= \delta_i \log(\lambda_{ij(i)}) - \sum_{j=1}^{j(i)} \lambda_{ij} t_{ij} && (1.17)\\
&= \sum_{j=1}^{j(i)} \left(\delta_{ij} \log \lambda_{ij} - \lambda_{ij} t_{ij}\right), && (1.18)
\end{aligned}$$

where (1.17) follows from $S(t_i; \mathbf{x}_i) = \exp(-\Lambda(t_i; \mathbf{x}_i)) = \exp(-\sum_{j=1}^{j(i)} \lambda_{ij} t_{ij})$ and (1.18) from $\delta_i \lambda_{ij(i)} = \sum_{j=1}^{j(i)} \delta_{ij} \lambda_{ij}$ since $\delta_{ij} = 0 \ \forall j \neq j(i)$.

Now let $\delta_{ij} \overset{iid}{\sim} Po(\mu_{ij})$ with $\mu_{ij} = \lambda_{ij} t_{ij}$ and density $f(\delta_{ij}) = \mu_{ij}^{\delta_{ij}} / \delta_{ij}! \cdot \exp(-\mu_{ij})$, where the factorial can be ignored since $\delta_{ij} \in \{0, 1\}$. In this case, the (Poisson) log-likelihood contribution for observation unit i is given by

$$\begin{aligned}
\ell_i(\boldsymbol{\beta}) = \log\left(\prod_{j=1}^{j(i)} f(\delta_{ij})\right) &= \sum_{j=1}^{j(i)} \delta_{ij} \log(\mu_{ij}) - \mu_{ij} \\
&= \sum_{j=1}^{j(i)} \delta_{ij} \log(\lambda_{ij}) + \delta_{ij} \log(t_{ij}) - \lambda_{ij} t_{ij} && (1.19)
\end{aligned}$$

Thus, the Poisson log-likelihood (1.19) is proportional to (1.18) as $\delta_{ij} \log(t_{ij})$ is independent of the parameters of interest. Therefore, parameter estimates for $\boldsymbol{\beta}$ can be obtained by optimizing the Poisson likelihood (1.19).

For concrete application, the link between expectation and linear predictor (1.15) is given by

$$\log(\mu_{ij}) = \log(\lambda_{ij} t_{ij}) = \eta_{ij} = \log(\lambda_j) + \mathbf{x}_i' \boldsymbol{\beta} + \log(t_{ij}). \tag{1.20}$$

The log-baseline hazard $\log(\lambda_j)$ in (1.20) is usually absorbed into the linear predictor, such that $\eta_{ij} = \tilde{\mathbf{x}}_i' \tilde{\boldsymbol{\beta}} + o_{ij}$ with $\tilde{\mathbf{x}_i}' = (1, I(j = 2), \ldots, I(j = J), \mathbf{x}_i')$, $\tilde{\boldsymbol{\beta}} = (\beta_{01}, \ldots, \beta_{0J}, \boldsymbol{\beta}')'$ and $\log(t_{ij}) =: o_{ij}$

enters the linear predictor as an offset with coefficient 1.

Historically, these relationships were discovered not long after the influential publication by Cox (1972). For example, Kalbfleisch and Prentice (1973); Holford (1976) present models based on the partitioning of the follow up in discrete intervals and discuss the conditions under which the resulting models are equivalent to Cox's suggestion. Holford (1980) additionally showed that for grouped data with groups $g = 1, \ldots, G$, $T_g = \sum_{i=1}^{n_g} t_{ig}$, $\lambda_g = \exp(\mathbf{x}_g'\boldsymbol{\beta})$ the models (i) $n_g \sim Po(\lambda_g T_g)$, (ii) $t_{ig} \sim Exp(\lambda_g)$ and (iii) $(n_1, \ldots, n_G) \sim Multinom(P_1, \ldots, P_G, \sum_g n_g)$ with $P_g = \frac{T_g}{\sum_g T_g}\lambda_g$, produce equivalent estimates for $\boldsymbol{\beta}$, when their likelihoods are optimized. Laird and Olivier (1981) made the link between model (1.15) and the Poisson regression more explicit, while Friedman (1982) discusses the asymptotic properties of the estimates, when the true underlying hazard is not a step function. Whitehead (1980); Aitkin and Clayton (1980); Clayton (1983) discuss fitting such proportional hazards models from a more practical perspective using GLIM (a popular implementation for maximum likelihood estimation at the time). Whitehead (1980) discussed in more detail the (implicit) handling of ties and the conditions under which the Poisson regression approach is equivalent to Cox models. Aitkin et al. (1983) compares different modeling approaches, including PEM by application to heart transplant data. Guo (1993) demonstrated that the PEM can be used to estimate models for left-truncated time-to-event data without much additional effort. Specifically, he shows that the in the case of left-truncated data, the log-likelihood contribution (1.19) of subject i is given by

$$\ell_i = \sum_{j=s(i)}^{j(i)} \delta_{ij} \log(\lambda_{ij}) - \delta_{ij} \log(t_{ij}) - \lambda_{ij} t_{ij},$$

where $s(i)$ is the interval in which subject i is first observed. The estimation is performed as described in Section 1.3 after all intervals with $j < s(i)$ were removed from the data for all $i = 1, \ldots, n$. Karrison (1996) derives confidence intervals for median survival times estimated by a PEM.

Despite this early interest in the PEM and its successful applications, survival analysis was mostly dominated by the Cox model in research as well as practical application. Part of the reason for the popularity of the Cox model compared to the PEM (despite their equivalence under certain conditions) was its computational efficiency, especially for increasing J, as only the partial likelihood has to be optimized, and availability of the Cox routine in standard statistical software. Another problematic point regarding the PEM is the arbitrariness of the split point selection (Demarqui et al., 2008), such that too few split points lead to crude approximations while fine split point grids can make the estimates of interval specific baseline hazards λ_j unstable, with large differences between neighboring intervals. One solution to these problems is provided by the piece-wise exponential additive model, as discussed in Section 1.5.

1.5 Piece-wise exponential Additive Mixed Model

To avoid the problems discussed in the previous paragraph, researchers soon realized that the baseline hazard could be estimated semi-parametrically. For example, Efron (1988) discussed many different models for discrete (or discretized) time-to-event data, including the PEM, and suggested to represent the baseline hazard via cubic-linear splines. Similarly, Carstensen (2005) illustrates the estimation of PEM where the baseline hazard is represented by natural cubic splines. The general idea is to use some representation of the smooth log-baseline hazard that is linear in the coefficients, as in (1.5). In this case, the part of the model matrix representing the baseline hazard is reduced to $K < J$ basis functions $B_k(t_j)$ evaluated at time-points $t_j \in (\kappa_{j-1}, \kappa_j], j = 1, \ldots, J$, e.g., $t_j := \kappa_j$. Because of the smoothness assumption of spline estimates, the problem of large differences between *neighboring* baseline hazards λ_j, λ_{j+1} is also somewhat meliorated.

However, such an approach essentially shifts the problem from choosing J and the placement of κ_j to the problem of choosing K and the placement of respective knots for the construction of basis functions, which again affects the results and is essentially arbitrary. One possibility to deal with this problem is to estimate multiple models with different K and knot placement and select the best model according to some evaluation criterion, e.g., AIC (Akaike, 1974), Brier Score (Gerds and Schumacher, 2006) or similar. The disadvantage of such approaches is that many, potentially computationally costly, models have to be fitted and that, as with any model or variable selection, subsequent inference will be invalid unless adjusted for the selection procedure (Berk et al., 2013).

One approach to deal with these problems is to choose K relatively large and penalize the wiggliness of the estimate, e.g., by imposing a quadratic penalty on the basis coefficients. Such models are referred to as GAMMs and were discussed in Section 1.3. In the context of time-to-event analysis, Cai et al. (2002) were among the first to use a mixed model based approach to obtain smooth hazard rate estimates. They derive the respective REML estimates and asses the performance of the approach via simulation studies. Later, Cai and Betensky (2003) extended their approach to interval censored data. Kauermann (2005) discussed mixed model based P-Spline smoothing for the estimation of PEMs with smoothly time-varying effects $f(t, x)$. Lambert and Eilers (2005) proposed a Bayesian P-Spline approach for the estimation of hazard models with time-varying effects, utilizing the connection of PEMs and the Poisson GLMs. Rodríguez-Girondo et al. (2013) discuss model building strategies for nonproportional hazards models that can be estimated via Poisson GAMs, including stepwise model selection based on the conditional AIC, double penalty procedures (Marra and Wood, 2011) and model based boosting (Hothorn and Bühlmann, 2006). In Sennhenn-Reulen and Kneib (2016) PAMMs were extended to multi-state models and estimated via structured fusion lasso.

Argyropoulos and Unruh (2015) introduce a slightly modified procedure. They use the Gauss-Lobatto integration rule to determine the positioning of the interval split points $\kappa_{j(i)}$, conditional on the

desired precision of the approximation. Note that κ_j then becomes a function of i, because each individuals follow-up is split up into J intervals whose positions depend on the observed t_i. Although this differs from the approach pursued in this work, where intervals are equal across subjects, most of the concepts like the inclusion of random effects (frailty), time-varying effects different time-scales are relevant here as well. They also discuss the advantages of representing the model by a Poisson GAM and specifically its estimation using the **mgcv** package (Wood, 2017), including big data methods (Wood et al., 2015, 2017).

In this work PAMMs with log-hazard (1.21) are considered

$$\log(\lambda_i(t; \mathbf{x}_i)) = \beta_0 + f_0(t) + \sum_{p=1}^{P} f_p(t, x_{i,p}), \tag{1.21}$$

which is an extension of (1.20) in that the log-baseline hazard λ_j is represented by $\beta_0 + f_0(t)$, where $f_0(t)$ denotes the non-linear contribution to the log-baseline hazard at time t, and covariate effects of time-constant covariates are potentially smooth, potentially smoothly time-varying effects $f_p(t, x_{i,p}), p = 1, \ldots, P$. For the sake of notational convenience, linear effects and effects of categorical covariates are subsumed in the above notation. A tutorial style introduction to such models with a focus on practical examples and comparison to the Cox model is presented in Bender et al. (2018a). Although not a focus in this work, Gaussian random effects (or frailties as they are often referred to in the context of time-to-event data analysis) can also be integrated in this framework (Wood and Scheipl, 2017).

1.6 Cumulative Effects

In this work (1.21) is further extended to include cumulative effects that occur when multiple past observations of a time-dependent covariate (TDC) affect the hazard at time t. Let $\mathbf{z}$ a time-dependent covariate (exposure) observed at exposure time points $t_z \in \{t_1, \ldots, t_Q\}$ and $z(t_z)$ the observed covariate value at exposure time t_z. A general representation of a cumulative effect is given in (1.22)

$$g(\mathbf{z}, t) = \int_{\mathcal{T}(t)} h(t, t_z, z(t_z)) \mathrm{d}t_z, \tag{1.22}$$

Thus, the cumulative effect on the hazard at time t depends on

- the timing of the exposure t_z,
- the amount of exposure $z(t_z)$,
- the specification of the partial effect $h(t, t_z, z(t_z))$, which controls how exposure $z(t_z)$ affects the hazard at time t and

- the *lag-lead* window $\mathcal{T}(t)$ that specifies which observations of $\mathbf{z}$ will contribute to the log-hazard at time t.

Conceptually, the partial effect defines how each exposure affects the hazard at the time of interest t, while the lag-lead window controls how many of such partial effects are cumulated at t. In practice, however, a misspecification of the lag-lead window might also affect the estimation of the partial effects. For notational convenience, in the following all partial effect specifications will be denoted by $h()$, even if the number of arguments changes.

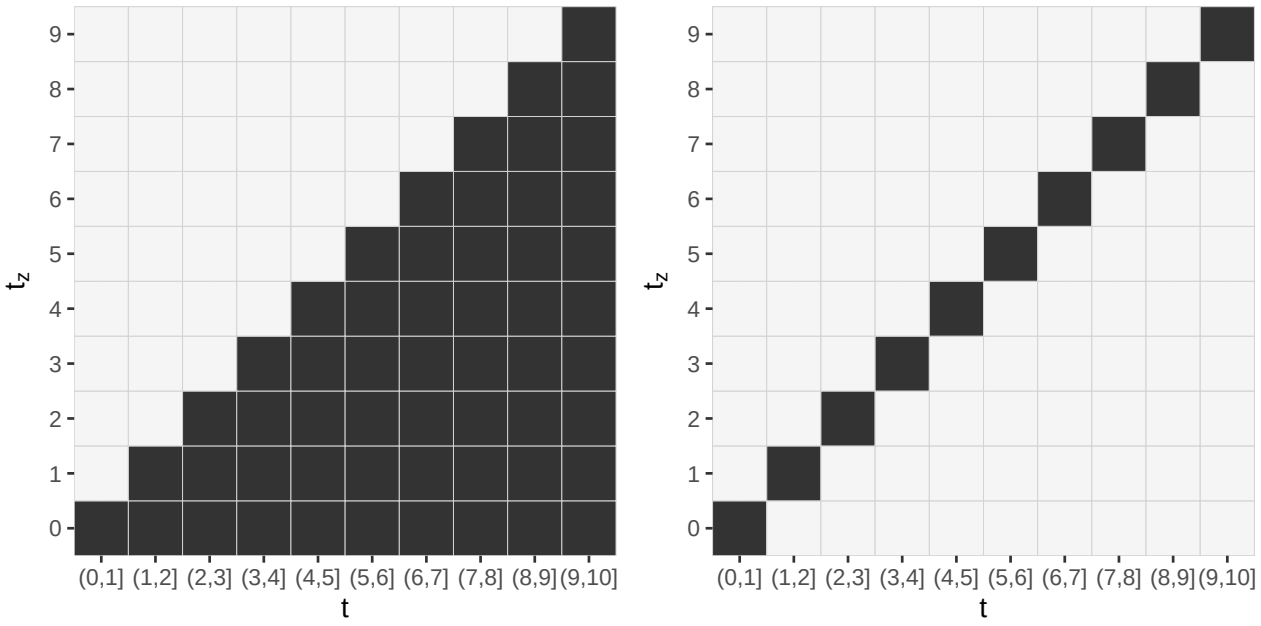


Figure 1.1: Visualization of two exemplary specifications of the lag-lead window $\mathcal{T}(t)$. The follow-up time was partitioned in 10 intervals. Exposures were observed at time-points $t_z = 0, \ldots, 9$. The lag-lead window on the left is defined by $\mathcal{T}(t) = \{t_{z,q} : t \geq t_{z,q}, q = 1, \ldots, 10\}$ and the lag-lead window on the right by $\mathcal{T}(t) = \{t_{z,q} : t = t_{z,q}, q = 1, \ldots, 10\}$. For the construction of such lag-lead windows, t should usually be the start time of the interval, otherwise an exposure observed at $t_z = 1$ would affect the hazard in interval $(0, 1]$.

Figure 1.1 shows two possible specifications of the lag-lead window $\mathcal{T}(t)$. For each interval on the x-axis, the black squares indicate the exposure time points t_z which contribute to the cumulative effect in the respective interval. The left panel of Figure 1.1 is equivalent to the minimal restriction that only past and current observations can contribute to the hazard at time t, the right panel is equivalent to the assumption that only the current observation of $z(t_z)$ contributes to the cumulative effect. In the latter case (1.22) can be simplified to $g(\mathbf{z}, t) = g(z(t), t) = h(t)z(t)$, which constitutes a so-called *concurrent* effect. If the weight function $h(t)$ is further assumed to be constant over time, this constitutes a simple linear effect of a TDC. Bender et al. (2018a) illustrate such an analysis in the context of PAMMs and compare the results to the original analysis in Fox and Weisberg (2011)

utilizing the extended Cox model.

First early work on such effects was provided by Vacek (1997) and Collet et al. (1999), who parameterized the effect by linear combinations of the form $\beta_1 z(t_1) + \beta_2 z(t_2) + \cdots$. This has the obvious disadvantage that effects of exposures $z(t_z)$ are assumed to be linear and independent between different time periods. Hauptmann et al. (2004) defines the cumulative effect as $g(\mathbf{z}, t) = \sum_{t_z=1}^{t} h(t_z) z(t_z)$, with weight function $h(t)$, which is represented basis expansion similar to (1.5) with B-Spline basis functions. This is a special case of (1.22), where the exposure is assumed to be observed on a regular grid (that coincides with the follow-up), that the partial effect is linear in z and non-linear over the exposure time. In the context of GAMMs such models are often referred to as *signal regression models* (Marx and Eilers, 2005). A similar model, embedded in the framework of Cox regression, suggested in Sylvestre and Abrahamowicz (2009) has gained wide popularity and is known as the *weighted cumulative exposure* model (WCE), defined in (1.23)

$$g(\mathbf{z}, t) = \sum_{t_z=1}^{t} h(t - t_z) z(t_z), \tag{1.23}$$

which is a reparametrization of Hauptmann et al. (2004)'s model, in that the effect of exposure at t_z does not depend on the exposure time itself, but rather on its *latency* $t - t_z$, i.e., the duration between the time at which the exposure was observed and the "current" follow-up time t. They represent the weight function $h(t - t_z)$ by cubic regression splines, while knot selection is performed using the Bayesian Information Criterion (Schwarz, 1978). They also discuss the possibility of constraining the weight function $h(t - t_z)$, such that it decreases to zero for increasing latency, by setting the two parameters associated with the two last basis functions to zero. Louveau et al. (2018) and Lèvèque et al. (2018) applied this approach to estimate the association between steroids and rheumatoid arthritis and smoking and exposure to asbestos on lung cancer, respectively. Mauff et al. (2017) extend the WCE approach to the context of Bayesian joint models, that are appropriate when the TDC is an internal variable and the survival process and the longitudinal process must be modeled simultaneously.

Berhane et al. (2008) were among the first who stated the general formulation (1.22), but only the bivariate relationship (1.24)

$$g(\mathbf{z}, t) = \int_0^t h(t - t_z, z(t_z)) \mathrm{d}t_z \tag{1.24}$$

is discussed explicitly, where $h(t - t_z, z(t_z))$ is represented by a tensor product via basis expansion (1.6) with B-Spline basis functions and the unpenalized estimation is performed by conditional logistic regression. In time-series analysis, models with cumulative effects (1.23) and (1.24) are referred to as *distributed lag (non-linear) models* (DL(N)M). Gasparrini et al. (2010) discusses such models extensively in the context of epidemiological studies and coined the term *exposure-lag-response associations* for such models (Gasparrini, 2014). Gasparrini et al. (2017) also discusses a penalized framework for the estimation of DLNMs. The **R** package **dlnm** implements these

methods (Gasparrini, 2011).

Conceptually, the PAMM with cumulative effects is very similar to historical effects (Malfait and Ramsay, 2008; Harezlak et al., 2007) in functional data analysis. In this context, Gellar et al. (2015) introduce Cox regression models with functional covariates and cumulative effects of the form $\int h(t_z)z(t_z)\mathrm{d}t_z$ that are estimated via penalized partial likelihood.

Usually, (1.22) is approximated by a weighted sum (e.g., trapezoidal rule)

$$g(\mathbf{z}, t) \approx \sum_{t_{z,q} \in \mathcal{T}(t)} \Delta_q h(t, t_{z,q}, z(t_{z,q})), \tag{1.25}$$

where Δ_q are quadrature weights $t_{z,q} - t_{z,q-1}$. For practical purposes this raises the question how to define Δ_1. Common choices are $\Delta_1 := \Delta_2$ and $\Delta_1 := \frac{1}{Q-1}\sum_{q=2}^{Q}\Delta_q$. Extending (1.6), the trivariate partial effect $h(t, t_z, z(t_z))$ can be represented by basis expansion (1.26)

$$h(t, t_{z,q}, z(t_{z,q})) = \sum_{\ell=1}^{L}\sum_{r=1}^{R}\sum_{m=1}^{M}\gamma_{\ell,r,m}B_m(z(t_{z,q}))B_r(t_{z,q})B_\ell(t), \tag{1.26}$$

with marginal design matrices $\mathbf{B}_t$, $\mathbf{B}_{t_z}$ and $\mathbf{B}_z$ analogous to Section 1.3 (cf. Wood, 2017, Ch. 5.6.1). The combined $n \times (L \cdot R \cdot M)$ design matrix for the cumulative effect (1.25) is then given by

$$\mathbf{B}_g = \mathbf{B}_t \otimes_r \mathbf{B}_{t_z} \otimes_r \mathbf{B}_z, \tag{1.27}$$

where $\otimes_r$ denotes the row-wise Kronecker product. For the estimation of the model with design matrix component (1.27), the respective entries where $t_{z,q} \notin \mathcal{T}(t)$ are set to zero and the entries for which $t_{z,q} \in \mathcal{T}(t)$ are multiplied by Δ_q.

The penalty matrix associated with (1.27) can also be constructed from marginal penalty matrices. Let $\mathbf{S}_z$, $\mathbf{S}_{t_z}$ and $\mathbf{S}_t$ the (suitably reparameterized; cf. (Wood, 2017, Ch. 5.6.2)) marginal matrices and $\tilde{\mathbf{S}}_z = \mathbf{I}_R \otimes \mathbf{I}_L \otimes \mathbf{S}_z$, $\tilde{\mathbf{S}}_{t_z} = \mathbf{I}_L \otimes \mathbf{S}_{t_z} \otimes \mathbf{I}_M$ and $\tilde{\mathbf{S}}_t = \mathbf{I}_R \otimes \mathbf{I}_L \otimes \mathbf{S}_z$, where $\mathbf{I}_R$, $\mathbf{I}_L$ and $\mathbf{I}_M$ are identity matrices of dimension R, L and M respectively and $\otimes$ denotes the Kronecker product. The total penalty associated with (1.27) is then defined by

$$\tilde{\mathbf{S}}_g = \nu_z\boldsymbol{\gamma}_g'\tilde{\mathbf{S}}_z\boldsymbol{\gamma}_g + \nu_{t_z}\boldsymbol{\gamma}_g'\tilde{\mathbf{S}}_{t_z}\boldsymbol{\gamma}_g + \nu_t\boldsymbol{\gamma}_g'\tilde{\mathbf{S}}_t\boldsymbol{\gamma}_g. \tag{1.28}$$

Notably, in (1.28), each margin is penalized by its own smoothing parameter (ν_z, ν_{t_z} and ν_t). This is an important feature of tensor product smooths (compared to multivariate isotropic smooths), as usually one would not expect the partial effects to be equally smooth with respect to the follow-up time, exposure time and covariate values. This allows the estimation of partial effects that, e.g., exhibit large (non-linear) variation over covariate values $z(t_z)$ but only moderately over t and t_z.

After the construction of design matrix $\mathbf{B}_g$ and penalty $\tilde{\mathbf{S}}_g$ as describe above, the coefficients $\boldsymbol{\gamma}_g$ associated with the cumulative effect can be estimated as described in Section 1.3. The general framework introduced in this section was developed in Bender et al. (2018b) and its application example discussed in more detail in Hartl et al. (2018). Note that in Bender et al. (2018b) the exposure times were denoted by t_e while in Bender and Scheipl (2018b) (and this work) the notation t_z is used to avoid confusion with **mgcv**'s function `te` that is used to estimate tensor product smooths.

1.7 Quantifying uncertainty

PAMMs model the (log-)hazard for which standard errors can be obtained straight forward from the theory of GAMMs (cf. Section 1.3). Since $\hat{\boldsymbol{\gamma}} \sim N(\boldsymbol{\gamma}, \mathbf{V}_{\boldsymbol{\gamma}})$, it follows that linear transformations of $\hat{\boldsymbol{\gamma}}$ also follow a normal distribution. For example, the (approximate) distribution of the log-hazard for covariate specification $\mathbf{x}$ is given by

$$\mathbf{x}'\hat{\boldsymbol{\gamma}} \sim N(\mathbf{x}'\boldsymbol{\gamma}, \mathbf{x}'\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{x}).$$

However, often we are also interested in quantities derived from the (log)-hazard, most notably the cumulative hazard $\Lambda(t|\mathbf{x})$ and the survival probability $S(t|\mathbf{x})$ that are non-linear transformations of $\boldsymbol{\gamma}$. Three common approaches to derive the standard errors or confidence intervals for such transformations are common in practice and in the literature:

1. Delta method (cf. Section 1.7.1)
2. Simulation based on the posterior distribution of the coefficients (cf. Section 1.2)
3. Direct transformation of the linear predictor (cf. Section 1.7.3)

In the following, the three approaches are briefly described in more detail. Results of a simulation study with respect to the coverage of the CI obtained by the different approaches, are presented in Section 1.8.

1.7.1 Delta method

Let $\hat{\boldsymbol{\gamma}}$ the $P \times 1$ vector of coefficient estimates and $h : \mathbb{R}^P \to \mathbb{R}^m$. Here m is the number of rows in $\mathbf{X}$. The Delta rule states, that the transformation $h(\hat{\boldsymbol{\gamma}})$ has normal distribution (1.29)

$$h(\hat{\boldsymbol{\gamma}}) \sim N\left(h(\boldsymbol{\gamma}), \mathbf{V}_{h(\boldsymbol{\gamma})} := \nabla h(\boldsymbol{\gamma})(\mathbf{V}_{\boldsymbol{\gamma}})\nabla h(\boldsymbol{\gamma})'\right), \tag{1.29}$$

where $\nabla h(\boldsymbol{\gamma})$ is the Jacobi matrix.

Below, the variances $\mathbf{V}_{h(\boldsymbol{\gamma})}$ are derived for the transformations

- $h(\boldsymbol{\gamma}) := \Lambda(t; \mathbf{X}, \boldsymbol{\gamma})$ and

- $h(\boldsymbol{\gamma}) := S(t; \mathbf{X}, \boldsymbol{\gamma}) = \exp(-\Lambda(t|\mathbf{X}, \boldsymbol{\gamma}))$.

In the following, let $t \equiv \kappa_J$ without loss of generality and $\mathbf{X}$ the $J \times P$ design matrix, such that the log-hazard in intervals $j = 1, \ldots, J$ is given by $\boldsymbol{\eta} = \mathbf{x}_1'\boldsymbol{\gamma}, \ldots, \mathbf{x}_J'\boldsymbol{\gamma} = \eta_1, \ldots \eta_J$. Let further $\mathbf{e}^{\mathbf{v}} := (\exp(v_1), \ldots, \exp(v_J))'$ and $\mathbf{E}^{\mathbf{v}}$ the respective diagonal matrix with elements $\mathbf{E}^{\mathbf{v}}_{j,j} = \exp(v_j), j = 1, \ldots, J$ and $\mathbf{E}^{\mathbf{v}}_{j,j'} = 0 \; \forall j' \neq j$.

Cumulative hazard

In the context of PAMMs, the cumulative hazard at time t is given by $\Lambda(t|\mathbf{X}, \boldsymbol{\gamma}) = \sum_{j=1}^{J} \Delta_j \exp(\eta_j)$, with $\Delta_j = \kappa_j - \kappa_{j-1}$, the length of the j-th interval. In matrix notation this can be written as $\boldsymbol{\Delta}\mathbf{e}^{\boldsymbol{\eta}}$, where $\boldsymbol{\Delta}$ is the lower triangular matrix (1.30)

$$\boldsymbol{\Delta} = \begin{pmatrix} \Delta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \Delta_1 & \cdots & \Delta_J \end{pmatrix} \tag{1.30}$$

These are known constants, thus it suffices to derive the Jacobi matrix for $\mathbf{e}^{\boldsymbol{\eta}}$, given in (1.33). Then

$$\nabla \mathbf{e}^{\boldsymbol{\eta}} = \begin{pmatrix} \partial \frac{\exp(\mathbf{x}_1'\boldsymbol{\gamma})}{\partial \gamma_1} & \cdots & \frac{\exp(\mathbf{x}_1'\boldsymbol{\gamma})}{\partial \gamma_P} \\ \vdots & \ddots & \vdots \\ \partial \frac{\exp(\mathbf{x}_J'\boldsymbol{\gamma})}{\partial \gamma_1} & \cdots & \frac{\exp(\mathbf{x}_J'\boldsymbol{\gamma})}{\partial \gamma_P} \end{pmatrix} \tag{1.31}$$

$$= \begin{pmatrix} \exp(\eta_1) \cdot x_{1,1} & \cdots & \exp(\eta_1) \cdot x_{1,P} \\ \vdots & \ddots & \vdots \\ \exp(\eta_J) \cdot x_{J,1} & \cdots & \exp(\eta_J) \cdot x_{J,P} \end{pmatrix} \tag{1.32}$$

$$= \mathbf{E}^{\boldsymbol{\eta}}\mathbf{X} \tag{1.33}$$

Thus, with (1.29) and (1.33), the variance of the cumulative hazard is given by

$$\text{Var}(\Lambda(t|\mathbf{X}, \boldsymbol{\gamma})) = (\boldsymbol{\Delta}\mathbf{E}^{\boldsymbol{\eta}}\mathbf{X})\mathbf{V}_{\boldsymbol{\gamma}}(\boldsymbol{\Delta}\mathbf{E}^{\boldsymbol{\eta}}\mathbf{X})' \tag{1.34}$$

$$= (\boldsymbol{\Delta}\mathbf{E}^{\boldsymbol{\eta}})(\mathbf{X}\mathbf{V}_{\boldsymbol{\gamma}}\mathbf{X}')(\boldsymbol{\Delta}\mathbf{E}^{\boldsymbol{\eta}})' \tag{1.35}$$

This result was also stated by Carstensen (2005) in a slightly less general form. Result (1.35) can be useful, when the variance of the linear transformation $\mathbf{X}\boldsymbol{\gamma}$ was already calculated in a previous step. When $t \neq \kappa_J$, $\mathbf{X}$ is a $j(t) \times P$ matrix and $\boldsymbol{\Delta}$ is a $j(t) \times j(t)$ matrix with elements $\Delta_1, \ldots, \Delta_{j(t)}$, where $j(t) = j : t \in (\kappa_{j-1}, \kappa_j]$ and $\Delta_{j(t)} = t - \kappa_{j(t)-1}$.

Survival probabilities

Results for the survivor function can be obtained similar to the cumulative hazard by defining $h(\boldsymbol{\gamma}) := S(t|\mathbf{X}, \boldsymbol{\gamma}) = \exp(-\Lambda(t|\mathbf{X}, \boldsymbol{\gamma}))$. The Jacobi matrix is then given by

$$\begin{aligned}
\nabla h(\boldsymbol{\gamma}) &= \begin{pmatrix} \frac{\partial e^{-\Delta_1 e^{\eta_1}}}{\partial \gamma_1} & \cdots & \frac{\partial e^{-\Delta_1 e^{\eta_1}}}{\partial \gamma_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial e^{-\sum_{j=1}^J \Delta_j e^{\eta_j}}}{\partial \gamma_1} & \cdots & \frac{\partial\ e^{-\sum_{j=1}^J \Delta_j e^{\eta_j}}}{\partial \gamma_P} \end{pmatrix} \\
&= \begin{pmatrix} e^{-\Delta_1 e^{\eta_1}} \cdot (-\Delta_1 e^{\eta_1}) \cdot x_{1,1} & \cdots & e^{-\Delta_1 e^{\eta_1}} \cdot (-\Delta_1 e^{\eta_1}) \cdot x_{1,P} \\ \vdots & \ddots & \vdots \\ e^{-\sum_{j=1}^J \Delta_j e^{\eta_j}} \cdot (-\sum_{j=1}^J \Delta_j e^{\eta_j} \cdot x_{j,1}) & \cdots & e^{-\sum_{j=1}^J \Delta_j e^{\eta_j}} \cdot (-\sum_{j=1}^J \Delta_j e^{\eta_j} \cdot x_{j,P}) \end{pmatrix} \qquad (1.36) \\
&= -\mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}} \boldsymbol{\Delta} \mathbf{E}^{\boldsymbol{\eta}} \mathbf{X} \qquad (1.37)
\end{aligned}$$

The variance of the survival probability follows with (1.29) and (1.37) as given below.

$$\begin{aligned}
\operatorname{Var}(S(t|\mathbf{X}, \boldsymbol{\gamma})) &= (-\mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}} \boldsymbol{\Delta} \mathbf{E}^{\boldsymbol{\eta}} \mathbf{X}) \mathbf{V}_{\boldsymbol{\gamma}} (-\mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}} \boldsymbol{\Delta} \mathbf{E}^{\boldsymbol{\eta}} \mathbf{X})' \qquad (1.38) \\
&= (-\mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}} \boldsymbol{\Delta} \mathbf{E}^{\boldsymbol{\eta}}) (\mathbf{X} \mathbf{V}_{\boldsymbol{\gamma}} \mathbf{X}') (-\mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}} \boldsymbol{\Delta} \mathbf{E}^{\boldsymbol{\eta}})' \qquad (1.39) \\
&= \mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}} \operatorname{Var}(\Lambda(t|\mathbf{X}, \boldsymbol{\gamma})) (\mathbf{E}^{-\boldsymbol{\Delta} \mathbf{e}^{\boldsymbol{\eta}}})'. \qquad (1.40)
\end{aligned}$$

Similar to (1.35), expression (1.39) can be used when the variance for $\mathbf{X}\boldsymbol{\gamma}$ is already available. Expression (1.40) can be useful, when the variance of the cumulative hazard was obtained in a previous calculation.

Practical considerations

If one is only interested in the square root of the diagonal elements of the variance formulas (1.34) and (1.38), i.e., the standard errors of the transformation $h(\boldsymbol{\gamma})$, the matrix $\mathbf{V}_{h(\boldsymbol{\gamma})} = \nabla h(\boldsymbol{\gamma}) \mathbf{V}_{\boldsymbol{\gamma}} \nabla h(\boldsymbol{\gamma})'$ doesn't have to be fully formed. Instead, the computationally more efficient procedure is to calculate $\tilde{\mathbf{V}}_{h(\boldsymbol{\gamma})} = (\nabla h(\boldsymbol{\gamma}) \mathbf{V}_{\boldsymbol{\gamma}}) \circ \nabla h(\boldsymbol{\gamma})$, with rows $\tilde{\mathbf{v}}_j = (\tilde{v}_{j,1}, \ldots, \tilde{v}_{j,J})$ and obtain standard errors from $se_j = \sqrt{\sum_{k=1}^J \tilde{v}_{j,k}}$ (e.g., Wood, 2017, p. 294).

1.7.2 Simulation based inference

When the estimation process of a GAMM is viewed from an empirical Bayes point of view, which, from a computational perspective, is equivalent to the REML based approach described in Section 1.3, it can be shown (e.g., Fahrmeir et al. (2013, Ch. 7.6.1), Wood (2017, Ch. 4.2.4, 5.8, 6.10)), that the posterior distribution of regression parameters $\boldsymbol{\gamma}$ is given by

$$\boldsymbol{\gamma}|\mathbf{y}, \boldsymbol{\nu} \sim N(\hat{\boldsymbol{\gamma}}, \mathbf{V}_{\boldsymbol{\gamma}}) \qquad (1.41)$$

with $\hat{\boldsymbol{\gamma}}$ and $\mathbf{V}_{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{S}_\nu)^{-1}$ as before. This result can be used to compute Bayesian credible intervals for any quantity of interest that is a function of regression coefficients $\boldsymbol{\gamma}$. In the context of PAMMs, this approach was described by Argyropoulos and Unruh (2015). For example, 95% CIs for $\hat{S}(t|\mathbf{x}_j)$ are obtained by drawing samples $\boldsymbol{\gamma}_r, r = 1, \ldots, R$ from (1.41), calculating $\hat{S}_r(t, \mathbf{x}_j)$. The lower and upper boarders of the CI is then obtained as the 2.5% and 97.5% quantiles of the R survival probabilities, respectively.

1.7.3 Direct transformation

One simple approach, which is often used in practice to calculating confidence intervals (CI) for monotone transformations of the linear predictor, is to apply the transformation to the lower and upper bound of the CI for the linear predictor. Thus, when $\hat{\eta}_j = \mathbf{x}_j'\hat{\boldsymbol{\gamma}}$ is the log-hazard in the j-th interval with CI $[\hat{l}_j = \hat{\eta}_j - \zeta_{1-\alpha/2}\hat{\sigma}_{\hat{\eta}_j}, \hat{u}_j = \hat{\eta}_j + \zeta_{1-\alpha/2}\hat{\sigma}_{\hat{\eta}_j}]$, with $\zeta_{1-\alpha/2}$ the $1-\alpha/2$ quantile of the normal distribution, the CI for $h(\hat{\eta}_j)$ is obtained by $[h(\hat{l}_j); h(\hat{u}_j)]$.

1.7.4 Summary

Section 1.8 demonstrates that all three approaches yield reasonable coverage when used to calculate CIs for the hazard, cumulative hazard or survival probability. The delta method produces symmetrical CIs, which is not always desirable, as it can lead to negative CIs for quantities that can not be negative by definition. The direct approach has good coverage properties (cf. Section 1.8.2) and has the advantage of being computationally efficient and that restrictions with respect to the support of the quantity of interest (e.g., $\lambda(t, \mathbf{x}, \boldsymbol{\gamma}) \geq 0 \ \forall t$) are automatically preserved, in case the restriction is implicit in the transformation function, e.g., $h(\boldsymbol{\gamma}) = \mathbf{e}^{\boldsymbol{\eta}} \geq 0$. However, it can only be used for monotonous transformations $h(\boldsymbol{\gamma})$ but not for example for cumulative hazard differences $\Lambda(t; \mathbf{X}_1) - \Lambda(t; \mathbf{X}_2)$. The latter is used in Bender and Scheipl (2018b) for comparison to the Aalen model Martinussen and Scheike (2006), in which case simulation based CI calculation is the easiest choice as the calculation of the delta method becomes increasingly tedious and depends on the specific covariate specification $\mathbf{X}_1$ and $\mathbf{X}_2$.

1.8 Empirical results

This section provides some empirical evaluations of the PAMM (1.21), specifically its sensitivity to different definitions of follow-up split points κ_j (cf. Section 1.8.1) and the coverage of CIs (cf. Section 1.8.2) obtained by the different methods described in Section 1.7. These simulations are also performed to establish the respective default settings for their implementation in **pammtools** (Bender and Scheipl, 2018a).

To this end, data with hazard

$$\lambda(t; x_1, x_2) = \exp(-3.5 + 6 \cdot f_0(t) - 0.5x_1 + \sqrt{x_2}), \tag{1.42}$$

where f_0 is the density of the $Gamma(2, 8)$ distribution, is simulated. Censoring times were drawn from a uniform distribution. Additionally, all subjects still alive at $\kappa_J = 10$ were censored at that time. Altogether this resulted in about 20% of observations being censored.

1.8.1 Placement of split points

For the first analysis the model was estimated with three different placements of follow-up split points:

- "default": split points at the unique observed event times
- "fine": a fine raster of equidistant split points with interval length 0.05
- "rough": a rough raster of equidistant split points with interval length 0.5

For each placement of split points, $R = 200$ data sets were simulated with $n = 500$ observations each and the hazard (1.2), cumulative hazard (1.3) and the survival probability (1.1) were estimated. The simulation runs were evaluated by the root mean squared error (RMSE)

$$RMSE = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{J_r} \sum_{j=1}^{J_r} h(\eta_{r,j}) - h(\hat{\eta}_{r,j}) \tag{1.43}$$

and coverage

$$C_\alpha = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{J_r} \sum_{j=1}^{J_r} I\{h(\gamma) \in [h(\hat{\eta}_{r,j}) \pm \zeta_{1-\alpha/2} \hat{se}_{h(\hat{\eta}_{r,j})}]\} \tag{1.44}$$

In (1.43) $\eta_{r,j}$ is the true linear predictor in the j-th interval based on the r-th simulated data set, evaluated at interval end-points κ_j and $\hat{\eta}_{r,j}$ the respective model estimate. J_r is the number of intervals for the r-th simulation run. For data simulation settings "fine" and "rough" $J_r = J \; \forall r = 1, \ldots, R$. The results are shown in Figure 1.2. In general, the PAMM estimates are relatively insensitive to the split point selection, although the average increases with decreasing J. Results presented in Table 1.1 also indicate that the "default" and "fine" split point selection schemes are preferable with respect to Coverage and RMSE.

1.8.2 Confidence Intervals

In this section the performance of the different methods to calculate confidence intervals discussed in Section 1.7 is evaluated in a small simulation study. The setup and evaluation are equivalent to the one in Section 1.8.1, however, only the "default" split point setting is considered. For each evaluation, the coverage of the CIs calculated by the Delta method (cf. Section 1.7.1), the simulation

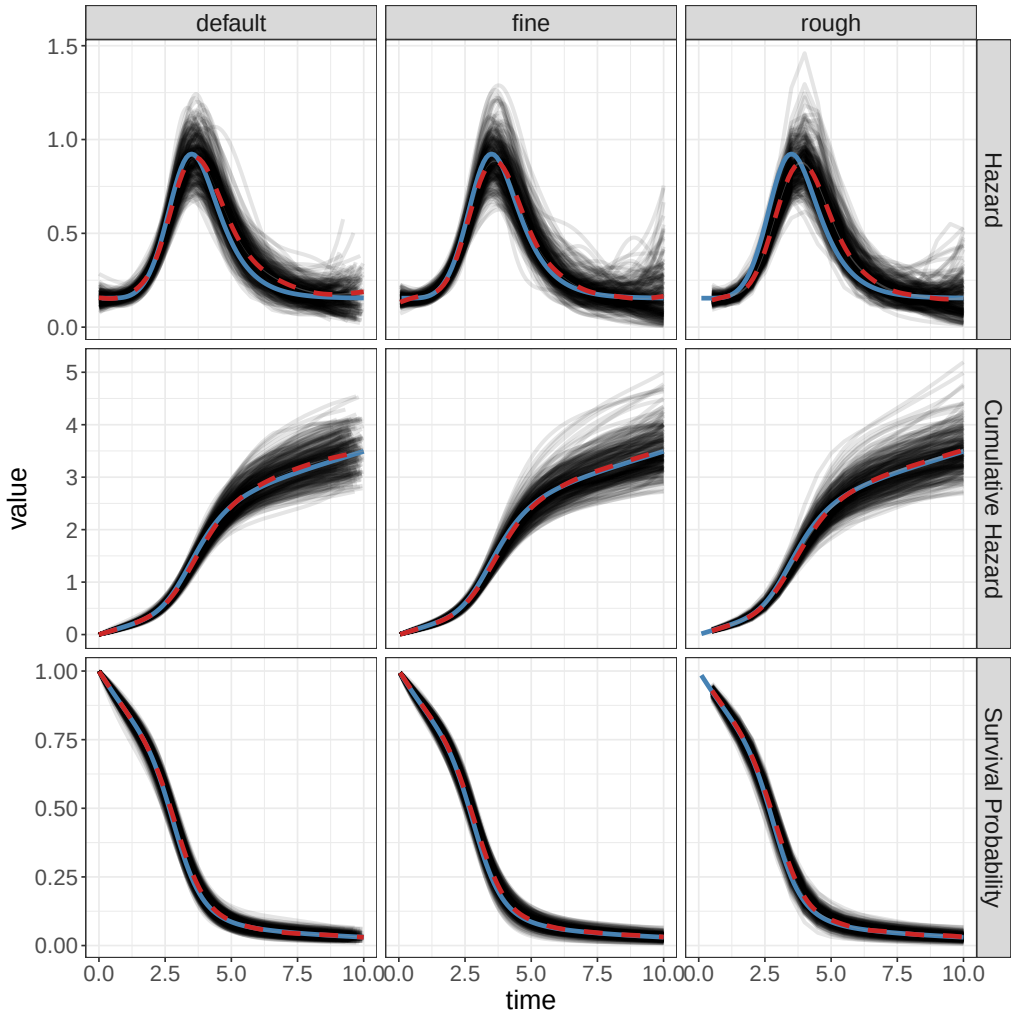


Figure 1.2: Simulation results for different placements of interval split points κ_j. Columns indicate the splitting scheme, rows indicate the evaluated quantity. Individual estimates are shown by black lines in the background, the solid blue line corresponds to the true underlying hazard and the dashed line the average estimate over all simulation runs. Note that for convenience the hazard estimates were displayed as lines instead of as step functions.

based approach (cf. Section 1.2) and the direct approach (cf. Section 1.7.3), respectively. A PAMM, specified as in Section 1.1 is fit to each data set. For each fit, the coverage of the CIs for the estimated hazard, cumulative hazard and survival probabilities was assessed.

The results are presented in Table 1.2. In general, the CIs are close to the nominal level of 95%. The Delta method and simulation based approaches show slight undercoverage for the cumulative hazard and the survival probability CIs, while the direct method shows slight overcoverage. The observed

		split points		
		default	fine	rough
Coverage	Hazard	0.92299	0.93250	0.82640
	Cumulative Hazard	0.97173	0.98484	0.91480
	Survival Probability	0.97173	0.98484	0.91480
RMSE	Hazard	0.06796	0.07328	0.10005
	Cumulative Hazard	0.09469	0.20181	0.23851
	Survival Probability	0.02269	0.01947	0.03317

Table 1.1: Coverage and RMSE for different split point placements. For the confidence intervals based on direct transformation (cf. Section 1.7.3), the coverage of the cumulative hazard and the survival probability is identical by definition.

	Method		
	Delta method	Simulation	Direct transformation
Hazard	0.9156	0.9010	0.9186
Cumulative Hazard	0.9175	0.9111	0.9736
Survival Probability	0.9231	0.9118	0.9736

Table 1.2: Estimated coverage C_α for the three confidence interval estimation methods introduced in Section 1.7. The simulation based estimation was based on 500 draws from the posterior of the model coefficients.

undercoverage for the hazard across CI calculation methods likely results from the hazard estimation being slightly biased (cf. upper left panel in Figure 1.2).

1.8.3 Conclusion

In conclusion, the simulation studies above suggest that the usage of split points at unique observed event times appears to be a sensible default value in settings with moderate censoring and medium to large number of observations (here $n = 500$). This is also confirmed by many applications to real data and comparison to other methods (e.g., Bender et al. (2018a); Bender and Scheipl (2018b)). Under the same circumstances, the direct transformation used to obtain CIs for the hazard, cumulative hazard and survival probability are performing reasonably well. Based on these results, the respective defaults used during data transformation and CI calculation in **pammtools** (Bender and Scheipl, 2018a) are set to the selection of split points at unique event times and the CI calculation by direct transformation. However, the other methods for CI calculation introduced here are also available.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6(4):701–726.

Aitkin, M. and Clayton, D. (1980). The fitting of Exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2):156–163.

Aitkin, M., Laird, N., and Francis, B. (1983). A reanalysis of the stanford heart transplant data. *Journal of the American Statistical Association*, 78(382):264–274.

Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, AC-19:716–723.

Alberda, C., Gramlich, L., Jones, N., Jeejeebhoy, K., Day, A. G., Dhaliwal, R., and Heyland, D. K. (2009). The relationship between nutritional intake and clinical outcomes in critically ill patients: results of an international multicenter observational study. *Intensive Care Medicine*, 35(10):1728–1737.

Argyropoulos, C. and Unruh, M. L. (2015). Analysis of time to event outcomes in randomized controlled trials by generalized additive models. *PLoS ONE*, 10(4):e0123784.

Bender, A., Groll, A., and Scheipl, F. (2018a). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, page 1471082X17748083.

Bender, A. and Scheipl, F. (2018a). adibender/pammtools: pammtools v.0.1.0.

Bender, A. and Scheipl, F. (2018b). pammtools: Piece-wise exponential additive mixed modeling tools. *arXiv:1806.01042 [stat]*.

Bender, A., Scheipl, F., Hartl, W. H., Day, A. G., and Küchenhoff, H. (2018b). Penalized estimation of complex, non-linear exposure-lag-response associations. *Biostatistics*.

Berhane, K., Hauptmann, M., and Langholz, B. (2008). Using tensor product splines in modeling exposure-time-response relationships: Application to the colorado plateau uranium miners cohort. *Statistics in Medicine*, 27(26):5484–5496.

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837.

Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics*, 59(3):570–579.

Cai, T., Hyndman, R. J., and Wand, M. P. (2002). Mixed model-based hazard estimation. *Journal of Computational and Graphical Statistics*, 11(4):784–798.

Carstensen, B. (2005). Practical use of the lexis diagram in the computer age or: Who needs the cox model anyway.

Clayton, D. G. (1983). Fitting a general family of failure-time distributions using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 32(2):102–109.

Collet, J.-P., Sharpe, C., Belzile, E., Boivin, J.-F., Hanley, J., and Abenhaim, L. (1999). Colorectal cancer prevention by non-steroidal anti-inflammatory drugs: effects of dosage and timing. *British Journal of Cancer*, 81(1):62–68.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, (Vol. 34, No. 2.):187–220.

Demarqui, F. N., Loschi, R. H., and Colosimo, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis*, 14(3):333–356.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American Statistical Association*, 83(402):414–425.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b -splines and penalties. *Statistical Science*, 11(2):89–121.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer Science & Business Media.

Fox, J. and Weisberg, H. S. (2011). *An R Companion to Applied Regression*. Sage Pubn, Thousand Oaks, Calif, revised. edition.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, 10(1):101–113.

Gasparrini, A. (2011). Distributed lag linear and non-linear models in `R`: the package `dlnm`. *Journal of Statistical Software*, 43(8):1–20.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine*, 33(5):881–899.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234.

Gasparrini, A., Scheipl, F., Armstrong, B., and Kenward, M. G. (2017). A penalized framework for distributed lag non-linear models. *Biometrics*, 73:938–948.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2015). Cox regression models with functional covariates for survival data. *Statistical Modelling*, 15(3):256–278.

Gerds, T. A. and Schumacher, M. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040.

Guo, G. (1993). Event-history analysis for left-truncated data. *Sociological Methodology*, 23:217–243.

Harezlak, J., Coull, B. A., Laird, N. M., Magari, S. R., and Christiani, D. C. (2007). Penalized solutions to functional regression problems. *Computational Statistics & Data Analysis*, 51(10):4911–4925.

Hartl, W. H., Bender, A., Scheipl, F., Kuppinger, D., Day, A. G., and Küchenhoff, H. (2018). Calorie intake and short-term survival of critically ill patients. *Clinical Nutrition*.

Hauptmann, M., Wellmann, J., Lubin, J. H., Rosenberg, P. S., and Kreienbrock, L. (2004). Analysis of exposure-time-response relationships using a spline weight function. *Biometrics*, 56(4):1105–1108.

Heyland, D. K., Cahill, N., and Day, A. G. (2011). Optimal amount of calories for critically ill patients: Depends on how you slice the cake! *Critical care medicine*, 39(12):2619–2626.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics*, 32(3):587–597.

Holford, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, 36(2):299–305.

Hothorn, T. and Bühlmann, P. (2006). Model-based boosting in high dimensions. *Bioinformatics*, 22(22):2828–2829.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on cox's regression and life model. *Biometrika*, 60(2):267–278.

Karrison, T. (1996). Confidence intervals for median survival times under a piecewise exponential model with proportional hazards covariate effects. *Statistics in Medicine*, 15(2):171–182.

Kauermann, G. (2005). Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics & Data Analysis*, 49(1):169–186.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76(374):231–240.

Lambert, P. and Eilers, P. H. C. (2005). Bayesian proportional hazards model with time-varying regression coefficients: a penalized poisson regression approach. *Statistics in Medicine*, 24(24):3977–3989.

Lèvèque, E., Lacourt, A., Luce, D., Sylvestre, M.-P., Guènel, P., Stücker, I., and Leffondrè, K. (2018). Time-dependent effect of intensity of smoking and of occupational exposure to asbestos on the risk of lung cancer: results from the icare case-control study. *Occup Environ Med*, pages oemed–2017–104953.

Louveau, B., De Rycke, Y., Lafourcade, A., Saraux, A., Guillemin, F., Tubach, F., Fautrel, B., and Hajage, D. (2018). Effect of cumulative exposure to corticosteroid and dmard on radiographic progression in rheumatoid arthritis: results from the espoir cohort. *Rheumatology*.

Malfait, N. and Ramsay, J. O. (2008). The historical functional linear model. *Canadian Journal of Statistics*, 31(2):115–128.

Marra, G. and Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7):2372–2387.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York, N.Y, auflage: 2006 edition.

Marx, B. D. and Eilers, P. H. (2005). Multidimensional penalized signal regression. *Technometrics*, 47(1):13–22.

Mauff, K., Steyerberg, E. W., Nijpels, G., van der Heijden, A. A., and Rizopoulos, D. (2017). Extension of the association structure in joint models to include weighted cumulative effects. *Statistics in Medicine*, pages n/a–n/a.

Patel, J. J., Lemieux, M., McClave, S. A., Martindale, R. G., Hurt, R. T., and Heyland, D. K. (2017). Critical care nutrition support best practices: Key differences between canadian and american guidelines. *Nutrition in Clinical Practice*, 32(5):633–644.

Rodríguez-Girondo, M., Kneib, T., Cadarso-Suárez, C., and Abu-Assi, E. (2013). Model building in nonproportional hazard regression. *Statistics in Medicine*, 32(30):5301–5314.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Sennhenn-Reulen, H. and Kneib, T. (2016). Structured fusion lasso penalized multi-state models. *Statistics in Medicine*.

Sylvestre, M.-P. and Abrahamowicz, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine*, 28(27):3437–3453.

Vacek, P. M. (1997). Assessing the effect of intensity when exposure varies over time. *Statistics in Medicine*, 16(5):505–513.

Whitehead, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(3):268–275.

Wood, S. and Scheipl, F. (2017). *gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4'*. R package version 0.2-5.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/Crc Texts in Statistical Science, Boca Raton, 2 rev ed. edition.

Wood, S. N., Goude, Y., and Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1):139–155.

Wood, S. N., Li, Z., Shaddick, G., and Augustin, N. H. (2017). Generalized Additive Models for Gigadata: Modeling the U.K. Black Smoke Network Daily Data. *Journal of the American Statistical Association*, 112(519):1199–1210.

Part I

Piece-wise exponential Additive Mixed Models

Chapter 2

A generalized additive mixed model approach to time-to-event analysis

Chapter 2 presents an application-oriented introduction of a model class that permits the usage of generalized additive mixed models for time-to-event data analysis, which we call piece-wise exponential additive mixed models. Special attention was given to time-varying effects and concrete workflows demonstrated using real data examples.

Contributing article:

Bender, A., Groll, A., and Scheipl, F. (2018a). A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, page 1471082X17748083

Author contributions:

Andreas Bender prepared a first draft, including worked examples and prepared the **R**-Code and visualizations. Andreas Groll revised the manuscript and contributed to all sections of the article. Fabian Scheipl revised the final draft of the manuscript and also contributed to all sections. All authors proofread the manuscript.

Supplementary material available at:

`http://www.statmod.org/smij/Vol18/Iss3-4/Bender/Abstract.html`
`https://zenodo.org/record/1147058`

Statistical Modelling 2018; **18(3–4)**: 1–23

A generalized additive model approach to time-to-event analysis

Andreas Bender[1], Andreas Groll[2] and Fabian Scheipl[1]
[1]Department of Statistics, Ludwig-Maximilians-Universität, München, Germany.
[2]Chairs of Statistics and Econometrics, Georg-August-Universität Göttingen, Germany.

Abstract: This tutorial article demonstrates how time-to-event data can be modelled in a very flexible way by taking advantage of advanced inference methods that have recently been developed for generalized additive mixed models. In particular, we describe the necessary pre-processing steps for transforming such data into a suitable format and show how a variety of effects, including a smooth nonlinear baseline hazard, and potentially nonlinear and nonlinearly time-varying effects, can be estimated and interpreted. We also present useful graphical tools for model evaluation and interpretation of the estimated effects. Throughout, we demonstrate this approach using various application examples. The article is accompanied by a new **R**-package called `pammtools` implementing all of the tools described here.

Key words: Cox model, time-varying coefficients, piece-wise exponential model, penalization, survival analysis, splines

Recieved May 2017; **revised** September 2017; **accepted** September 2017

1 Introduction

In this tutorial, we introduce a general framework for fitting time-to-event data, that is, the time until an event occurs, denoted by (the random variable) T. Classical application examples, which we will discuss in more detail later, include time until death of cancer patients and time until convicts are rearrested. When modelling such data, the central property of interest is usually the survival function $S(t) := \mathbb{P}(T > t)$, that is, the probability for an event occurring after time t. The modelling of such data, however, generally focuses on the so-called hazard rate, in the following denoted by $\lambda(t)$, which represents the instantaneous (normalized) risk of having an event in t, given no event occurred up to time t. The corresponding mathematical, rather technical definition of the hazard rate, is given as follows:

$$\lambda(t) := \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{1.1}$$

Address for correspondence: Andreas Groll, Chairs of Statistics and Econometrics, Georg-August-Universität, Humboldtallee 3, 37073 Göttingen, Germany.
E-mail: agroll@uni-goettingen.de

 10.1177/1471082X17748083

In the following, we demonstrate how a large class of models for time-to-event data can be represented as generalized additive mixed models (GAMMs). Using this representation, which requires a specific transformation of the original time-to-event data (see Section 2.2 for details), all the methods and versatile software implementations available for GAMMs, many of which are covered in this special issue, can be applied to survival analysis. In this way, the specification and penalized estimation of, for example, nonlinear, spatial or random effects for time-to-event data becomes routine and rather easy to do, equally so for (nonlinearly) time-varying effects.

The representation described here is not novel and known in the literature as piece-wise exponential models (PEMs). Under certain assumptions, PEMs are essentially Poisson generalized linear models (GLMs; Holford, 1980; Laird and Olivier, 1981; Friedman, 1982) with likelihoods proportional to the (partial) likelihood of a corresponding Cox model (see Cox, 1972 and Equation (2.1)). The PEM representation was popular temporarily when implementations of GLMs were more readily available in different statistical software packages compared to implementations of dedicated algorithms for survival models (Whitehead, 1980; Clayton, 1983), but most research in the field of survival analysis has concentrated on the Cox model and its extensions or on models based on the counting process representation of time-to-event data (Andersen et al., 1992; Martinussen and Scheike, 2006).

The PEM representation requires a partition of the follow-up time into a finite number of intervals and assumes that hazard rates are piece-wise constant in each of these intervals. The arbitrary choice of cut-points defining this partition has often been a point of criticism regarding PEMs. If the number of cut-points is too small, the step function approximation of the hazard rate may be too crude. A large number of cut-points, on the other hand, may lead to over-fitting and inefficient estimation with unstable estimates, as the baseline hazard requires the estimation of one parameter per interval. We are convinced, however, that the usage of PEMs remains desirable, especially in situations where one wants to take advantage of the methodological (Wood, 2011) and algorithmic (Wood, 2017) advances that have been made for GAMMs over the last years, particularly whenever inclusion of nonlinear, multivariate, spatial or spatio-temporal, or random effects is required. Even more importantly, the current state of the art for additive models allows analysts to largely avoid the 'arbitrary cut-points' problem of classical PEMs. Analysts can simply use a large number of cut-points and estimate the baseline hazard and other time-varying effects semiparametrically, while avoiding over-fitting and instability by means of penalization. We call this extension of the PEM a Piece-wise exponential Additive Mixed Model (PAMM).

This idea has been utilized in many recent publications. For example, Rodríguez-Girondo et al. (2013) use this approach to discuss model building strategies, including double shrinkage methods. Argyropoulos and Unruh (2015) discuss a variant of this method, which they call Poisson generalized additive model (GAM) using a Gauß–Lobato quadrature rule to partition the follow-up. They also give a thorough overview of methods for flexible parametric inference for time-to-event data as well as an overview of previous research on the Poisson

model for survival analysis. Sennhenn-Reulen and Kneib (2016) use PAMMs to fit LASSO-penalized multi-state models, while Gasparrini et al. (2017) extend distributed lag nonlinear models known from time-series analysis to the analysis of time-to-event data. Despite these recent publications and applications, the general idea of using GAMMs in the context of survival analysis is still not widely known by practitioners in substantive sciences, and the application of PAMMs is hindered by the lack of readily available clear-cut instructions on practical details such as data transformation and dedicated software that facilitate preparation and evaluation of such models.

Thus, the goal of this tutorial is to introduce and describe the general idea of the classical PEM as well as its semiparametric extension, the PAMM, and illustrate their use, starting from data transformation and standard models to more advanced applications. This tutorial is aimed at practitioners and focuses on building intuition and providing applicable advice rather than methodological detail and mathematical rigour. References for further reading are provided throughout. All results presented in this tutorial can be reproduced using the instructions and code in the vignettes of the add-on package `pammtools` (Bender and Scheipl, 2017), as described in Section 4.

In the following sections, we give a brief introduction to the PEM (Section 2) and its mathematical representation as a Poisson GLM (Section 2.1), as well as the data transformation required for its estimation (Section 2.2). We then describe the transition from PEMs to PAMMs in Section 2.3 and briefly outline the advantages of this approach. In Section 3, several examples for the application of PAMMs are provided, ranging from very simple (baseline hazard, time-constant effects) to more advanced models (stratified baseline hazards, nonlinearly time-varying effects, etc.).

2 Piece-wise exponential (additive) models

PEMs represent an alternative to classical approaches for continuous time-to-event data like, for example, the Cox model. If the partition of the follow-up time is selected with care, PEMs allow analysts to take advantage of all of the methodological and algorithmic advances that have been made for GAMMs over the last decades. In particular, it is fairly easy to include diverse types of effects such as nonlinear, multivariate, spatial, spatio-temporal or random effects in existing software packages for GAMMs.

Figure 1 illustrates the basic idea of a PEM for time-to-event data by applying it to survival times drawn from a Weibull distribution. To estimate the true underlying Weibull hazard rate (left panel), the follow-up is partitioned into a fixed number of intervals (here $J = 5$) with interval cut-points $\kappa_0 = 0 < \cdots < \kappa_J = 4$ (mid panel) and a constant hazard is estimated for each interval (right panel). Thus, the name piece-wise *exponential*—the hazard rate of an exponential distribution is constant over time. The approximation in Figure 1 appears rather crude, but given enough cut-points, PEM and PAMM estimates are very similar (or even equivalent) to Cox regression estimates, as demonstrated in Section 3.

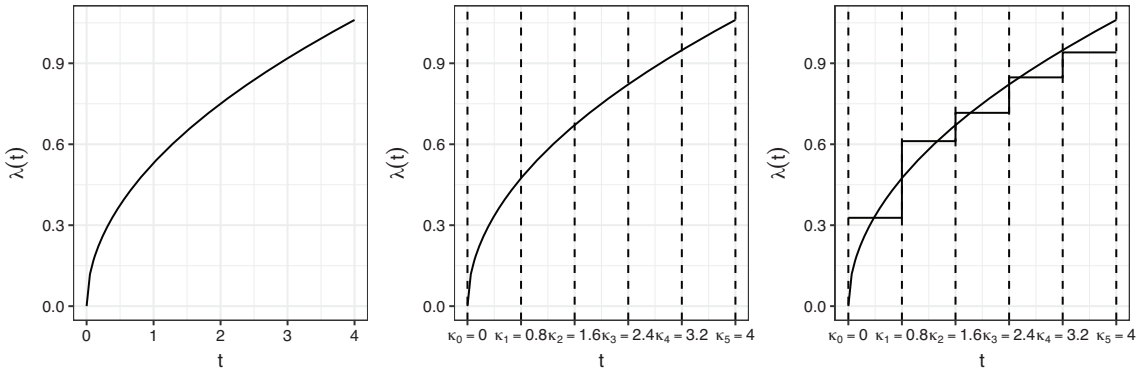


Figure 1 Hazard rate of a Weibull distribution (left panel); partitioning of the follow-up into $J = 5$ intervals (mid panel); estimate of the hazard rate via interval-specific piece-wise constant hazards, obtained by fitting a PEM to the data (right panel)

The difference between a PEM and a PAMM then results from the different approaches for the estimation of the baseline hazard and other smooth, potentially time-varying effects. It is important to note that although PAMMs model the baseline hazard using a smooth, nonlinear function, the resulting (estimated) hazard rate is still piece-wise constant (see Section 2.3).

2.1 The Poisson-likelihood of a PEM

Following Bender et al. (2016), we now introduce PEMs more formally. We start with a general proportional hazards (PH) model given by

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \qquad i = 1, \ldots, n, \tag{2.1}$$

where n is the number of subjects under study, $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\top$ is the row vector of time-constant covariates for subject i, and $\boldsymbol{\beta}$ the vector of corresponding regression coefficients. In the framework of the Cox PH model, the parameters $\boldsymbol{\beta}$ are estimated via partial likelihood, while the baseline hazard $\lambda_0(t)$ is estimated non-parametrically via the Nelson–Aaalen estimator. A PEM is obtained by partitioning the follow-up period $(0, t_{\max}]$, where $t_{\max}$ denotes the maximal (observed) follow-up time, into J intervals. To this end, we define $J + 1$ cut-points $0 = \kappa_0 < \cdots < \kappa_J = t_{\max}$. The j-th interval is then given by $(\kappa_{j-1}, \kappa_j]$. Assuming a constant hazard rate within each interval j, that is, $\lambda_0(t) = \lambda_j \; \forall t \in (\kappa_{j-1}, \kappa_j], t > 0$, (2.1) simplifies to

$$\lambda_i(t|\mathbf{x}_i) = \lambda_j \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad \forall\, t \in (\kappa_{j-1}, \kappa_j]. \tag{2.2}$$

If the underlying time-to-event data is structured in a certain way, that is, containing event indicators δ_{ij} and offsets o_{ij} for all intervals j in which subject i is under risk, it can be shown (Friedman, 1982) that the likelihood of the Poisson regression model

$$\mathbb{E}(\delta_{ij}|\mathbf{x}_i) = \exp(\log(\lambda_j) + \mathbf{x}_i^\top \boldsymbol{\beta} + o_{ij}) \tag{2.3}$$

is proportional to the one of model (2.2). Consequently, the two models are equivalent with respect to the ML estimator of $\boldsymbol{\beta}$. In practice, when fitting the respective Poisson GLM, $\log(\lambda_j)$ is incorporated in the linear predictor $\mathbf{x}_i^\top \boldsymbol{\beta}$, such that the design matrix $\mathbf{X}$ contains J additional dummy coded columns, where column j takes value 1 in rows representing observations in interval $(\kappa_{j-1}, \kappa_j]$ and 0 otherwise. The o_{ij} are simply added to the linear predictor as a so-called *offset*, which is usually of little interest, but note that the offset contains the information on the actual observed survival time in each interval, and thus makes the PEM a model for continuous time-to-event data, such that (2.3) can be reformulated as $\lambda_i(t|\mathbf{x}_i) = \dfrac{\mathbb{E}(\delta_{ij}|\mathbf{x}_i)}{t_{ij}} = \lambda_j \exp(\mathbf{x}_i'\boldsymbol{\beta})$ (where $t_{ij} = \exp(o_{ij})$; cf. Section 2.2). For the remainder of the article, we will omit the offset o_{ij} from the model specification, but it is important to stress that the offset must be included at the estimation stage.

Note that if the survival time is in fact discrete, was only observed on a discrete grid or can be reasonably discretized without loss of information, most of the methods and strategies presented here are equally applicable to discrete-time models. These are described in detail in the tutorial by Berger and Schmid (2018), which is also part of this special issue.

In the next section, we illustrate how to transform time-to-event data to be in accordance with the model specification (2.3).

2.2 Data transformation

To fit the PEM introduced in the previous section, time-to-event data needs to be transformed and restructured in a specific way. First, for each subject i let T_i denote its true survival time and C_i its (non-informative) censoring time. Then, $t_i := \min(T_i, C_i)$ represents the observed right-censored time under risk for subject i. Given intervals $1, \ldots, J$ and observed right-censored times t_i, for each time interval j that subject i is under risk, one has to create

(a) the binary response δ_{ij} as interval-specific event indicator, with $\delta_{ij} = 1$, if both, $\{t_i \in (\kappa_{j-1}, \kappa_j]\}$ and $\{t_i = T_i\}$, and $\delta_{ij} = 0$ else;
(b) the offset $o_{ij} = \log(t_{ij})$, based on the time subject i is under risk in interval j, which is given by $t_{ij} = \min(t_i - \kappa_{j-1}, \kappa_j - \kappa_{j-1})$.

We illustrate the data transformation here for the simple case of survival data without covariates. Consider, for instance, the survival times of the two subjects in Table 1. The same data in the format of piece-wise exponential data (PED) is shown in Table 2. Each subject has as many entries (rows) as the number of intervals for which the subject is included in the risk set. Note that these rows can then be treated as independent (given covariates) within the estimation scheme. The intervals $(\kappa_{j-1}, \kappa_j]$,

Table 1 Example of (conventional) time-to-event data for two subjects, $i \in \{1, 2\}$. Subject $i = 1$ is censored at $t_1 = 0.5$ and subject $i = 2$ experienced an event at $t_2 = 2.7$

i	t_i	δ_i
1	0.5	0
2	2.7	1

Table 2 Data in the piece-wise exponential format with one row per interval in which a subject was in the risk set. Intervals are defined by $J + 1$ cut-points $\kappa_0 = 0 < \cdots < \kappa_J = 4$; δ_{ij} is the status of subject i in interval j, t_{ij} the time subject i spent in interval j; the offset is denoted by $o_{ij} = \log(t_{ij})$

i	j	$(\kappa_{j-1}, \kappa_j]$	δ_{ij}	t_{ij}	$o_{ij} = \log(t_{ij})$
1	1	(0, 0.8]	0	0.5	$\log(0.5) = -0.69$
2	1	(0, 0.8]	0	0.8	$\log(0.8) = -0.22$
2	2	(0.8, 1.6]	0	0.8	$\log(0.8) = -0.22$
2	3	(1.6, 2.4]	0	0.8	$\log(0.8) = -0.22$
2	4	(2.4, 3.2]	1	0.3	$\log(0.3) = -1.20$

$j = 1, \ldots, J$ are specified by the user. Here, we chose equidistant intervals of length 0.8 as in Figure 1. Readers familiar with survival analysis will recognize this data structure to be very similar to the 'start-stop' format used to fit time-varying effects or effects of time-dependent covariates (TDCs) in the extended Cox regression model (Thomas and Reyes, 2014), with κ_{j-1} and κ_j as start and stop times, respectively. Further details and illustrations on transforming conventional time-to-event data into the format suitable to fit PEMs are provided in the data-transformation vignette (cf. Section 4 for details).

Note that another major advantage of this data structure is that (piece-wise constant) TDCs, that is, covariates that change their value over the follow-up period, can be incorporated naturally. In this case, the interval cut-points κ_j must additionally include all time points at which changes in the TDC are recorded and model (2.2) is extended to the form $\lambda_i(t|\mathbf{x}_{ij}) = \lambda_j \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta})$ (cf. Section 3.4). Furthermore, time-varying effects, where the association between a covariate and the hazard rate is allowed to change over the follow-up, can be incorporated by including an interaction term of the covariate with time t in the linear predictor (for more details, see Sections 3.3.2 and 3.3.3). This requires defining a TDC for time itself, for example, by setting $t_j := \kappa_j$ in the respective rows of the transformed data set.

Similarly, estimation of left-truncated data can be easily accommodated by excluding intervals before the respective left-truncation times as described by Guo (1993).

2.3 Piece-wise exponential Additive Mixed Model

Model (2.3) can be extended to include nonlinear or smoothly time-varying effects of time-constant or TDCs by incorporating semiparametric effects, that is, moving from a Poisson GLM representation to a Poisson GAMM representation. In reference to

the acronyms for PEMs and GAMs, we denote this model type by PAMM (piece-wise exponential additive mixed model). For a general PAMM, the hazard rate at time t for individual i with covariate vector $\mathbf{x}_i$ is given by

$$\lambda_i(t|\mathbf{x}_i) = \exp\left(f_0(t_j) + \sum_{k=1}^{p} f_k(x_{i,k}, t_j) + b_{\ell_i}\right), \quad \forall\, t \in (\kappa_{j-1}, \kappa_j]\,, \tag{2.4}$$

where $f_0(t_j)$ represents the log-baseline hazard rate (cf. Section 2.3.1) and $f_k(x_{i,k}, t_j)$, $k = 1, \ldots, p$, denotes very general effect types, possibly of different complexity and potentially depending on both a covariate and time. In particular, smooth nonlinear and smoothly time-varying effects (cf. Section 2.3.2) of the time-constant confounders $\mathbf{x}_{\bullet k} = (x_{1,k}, \ldots, x_{n,k})^{\mathrm{T}}$ are included. The time variable t_j is constant over each interval to ensure that the estimated hazard rates still correspond to a PEM. Typical choices are interval end-points $t_j = \kappa_j\, \forall\, t \in (\kappa_{j-1}, \kappa_j]$, or interval mid-points $t_j = \dfrac{\kappa_j + \kappa_{j-1}}{2}\, \forall\, t \in (\kappa_{j-1}, \kappa_j]$. Additionally, b_{ℓ_i} denote random intercept terms (Gaussian frailty), where ℓ_i, $\ell = 1, \ldots, L$ is the cluster to which subject i belongs. We do not discuss these terms in this tutorial, but an example application is provided in the `pammtools` vignette frailty (cf. Table 7).

A common way to specify unknown smooth functions like $f(x, t_j)$ is to use splines, which are represented by a weighted sum of M basis functions. A univariate function, such as the baseline hazard $f_0(t_j)$, can be expanded as $f_0(t_j) = \sum_{m=1}^{M} \gamma_{0m} B_m(t_j)$, with *spline coefficients* γ_{0m} and *basis functions* $B_m(t_j)$ (for a detailed description of splines, see, for example, Ruppert et al., 2003). Increasing the number of basis functions for such terms increases their flexibility and also the danger of overfitting, while an excessively low number of basis functions might not provide the necessary flexibility. In PAMMs, this trade-off is resolved by specifying a relatively large number of basis functions and then *penalizing* 'wiggliness' of the estimate (e.g., by penalizing the difference between neighbouring basis coefficients in P-splines; Eilers and Marx, 1996). Technically, the penalization strength is controlled by separate smoothing parameters for each additive term, which are estimated simultaneously with all other model coefficients, for example, using restricted maximum likelihood (REML) estimation (Wood, 2011). This has the advantage that no strong assumptions about the shapes of such smooth effects are necessary in order to specify the model. Instead, they are estimated based on the data.

This idea can be extended to bivariate interaction surfaces, such as the $f_k(x_{i,k}, t_j)$ terms in Equation (2.4), through the use of tensor product bases of the form $f(x_{i,k}, t_j) = \sum_{m=1}^{M} \sum_{\ell=1}^{L} \gamma_{m\ell} B_m(x_{i,k}) B_\ell(t_j)$. By specifying an interaction of covariates $\mathbf{x}_{\bullet k}$ with the (discretized) time t_j, we can model (piece-wise constant) time-varying effects of different grades of complexity, the most common of which are summarized in Table 3. Section 3.3 provides examples for most of these effects illustrated on widely known data sets. Note here that the effect types in Table 3 are merely a small subset of

Table 3 Overview of potentially smooth nonlinear and/or smoothly time-varying effect specifications in the analysis of time-to-event data

Effect Specification	Description
$\beta_k x_{i,k} + \beta_{k:t_j} \cdot x_{i,k} \cdot t_j$:	Linear, linearly time-varying effect
$f_k(x_{i,k})$:	Smooth nonlinear, time-constant effect
$f_k(x_{i,k}) \cdot t_j$:	Smooth, linearly time-varying effect
$x_{i,k} \cdot f_k(t_j)$:	Linear, smoothly time-varying effect
$f_k(x_{i,k}, t_j)$:	Smooth, smoothly time-varying effect

potential effect types. Further extensions such as spatial effects can be incorporated into Equation (2.4) just as easily, but they are beyond the scope of this tutorial.

Next (cf. Section 2.3.1), we illustrate the estimation of the baseline hazard with PEMs and PAMMs, and compare the results with the respective Nelson–Aalen estimates.

2.3.1 Baseline hazard

In the original definition of PEMs in (2.3), the baseline hazard is modelled by a step function with interval-specific hazards λ_j, which are estimated by including dummy variables for each interval in the design matrix. This has two major drawbacks:

1. the choice of the interval cut-points as well as the number of intervals is rather arbitrary (c.f. Demarqui et al., 2008); and
2. if a large number of cut-points is used, (too) many parameters need to be estimated and the individual estimates $\hat{\lambda}_j$ can become unstable. If the estimation of time-varying effects is required as well, this problem is exacerbated further.

In PAMMs, the baseline hazard is modelled as a regression spline over time, using a suitably discretized time variable t_j as earlier defined. Choosing a sufficiently large number of intervals and spline basis functions, the baseline hazard can then be estimated very flexibly. At the same time, over-fitting is avoided via penalization and, hence, smooth and stable estimates are obtained.

In many real life applications, the (baseline) hazard changes quickly at the beginning of the follow-up and less so towards the end. In these cases, a fixed penalty for the whole spline may be too restrictive; thus an estimation of the baseline hazard using adaptive spline smooths, where the penalty applied to the basis coefficients can itself change over the course of the follow-up (i.e., smaller penalty at the beginning, stronger penalty towards the end) may be preferable (Wood, 2011). This, however, also implies a higher computational burden.

2.3.2 Smooth nonlinear, smoothly time-varying effects

The major advantage of PAMMs over currently available implementations of Cox-type or Aalen-type models is that the entire flexibility and methodological progress of GAMMs can be employed with regard to covariate effects, not only to

the reliable and smooth estimation of baseline hazard rates. The summands $f_k(x_{i,k}, t_j)$ in the second term in Equation (2.4) can represent a variety of different effect types, ranging from simple linear, time-constant effects $x_{i,k}\beta_k$ to nonlinear, time-constant effects $f_k(x_{i,k})$ and nonlinear and smoothly time-varying effects $f_k(x_{i,k}, t_j)$. Table 3 summarizes the most common and important of these effect types. Time-varying effects are modelled as interaction terms between the covariate of interest, that is, $\mathbf{x}_{\bullet k}$, and the discretized time t_j, that is, the corresponding interval end or mid-points. All of these effect types can be specified fairly easily by the practitioner, for example, within the syntax of the `gam` function from the **R**-package `mgcv` (Wood, 2017), if the data is given in the format of Table 2 complemented by covariate information.

3 Applications and illustrations

In the following section, we describe and illustrate the application of PAMMs to various settings in time-to-event analysis and compare the estimates obtained from PAMMs with other established approaches on real data examples. For illustration, we use a follow-up restricted ($t \leq 400$) subset of the Veterans' Administration lung cancer study (Kalbfleisch and Prentice, 1980), which is available in the `survival` (Therneau, 2015) package and is a widely used data example in the context of time-to-event analysis. In this study, males with advanced inoperable lung cancer were randomized to one of two treatment regimens for lung cancer, either standard or novel chemotherapy, represented by the binary variable *trt*. In addition, the data set contains several time-constant covariates, namely *celltype* (categorical) of the tumour, the *age* (in years) at the beginning of treatment, the Karnofsky performance score (*karno*; 100 = good) and a dummy variable *prior*, indicating whether there had been a prior therapy (0 = no, 1 = yes) along with survival *time* and censoring *status* (0 = censored, 1 = event). Within the scope of this study, it is of major interest how the two different treatments in combination with the other covariates affect the survival of lung cancer patients. An excerpt of the data can be found in Table 4.

We omit the respective **R**-code here for clarity and brevity, however, each section refers to a dedicated vignette in the `pammtools` package that contains code showing the practical work flow in full detail (cf. Section 4).

3.1 Estimating the baseline hazard

First, we demonstrate how to fit and visualize simple baseline models using the `pammtools` package. We use both a PEM and a PAMM and compare the results of the estimated baseline hazard to the conventional Nelson–Aalen estimator, which can be obtained, for example, by applying the `coxph` function from the `survival` package directly on the `veteran` data from Table 4. In order to fit PEMs and PAMMs, however, we need to transform the data into the PED format, similar to the exemplary data presented in Table 2 using the `split_data` function from the `pammtools` package.

Table 4 Raw structure of the Veterans' Administration lung cancer study data

trt	celltype	time	status	karno	age	prior
1	large	19	1	30	39	1
1	squamous	231	0	50	52	1
0	large	156	1	70	66	0
0	smallcell	51	1	60	67	0
1	smallcell	95	1	70	61	0
1	large	133	1	75	65	0

Table 5 Veterans' Administration lung cancer study data after transformation to the piece-wise exponential data (PED) format (first six rows of subject 1)

id	tstart	tend	interval	offset	ped_status	trt	...	prior
1	0	1	(0,1]	0	0	1	...	1
1	1	2	(1,2]	0	0	1	...	1
1	2	3	(2,3]	0	0	1	...	1
1	3	4	(3,4]	0	0	1	...	1
1	4	7	(4,7]	1.1	0	1	...	1
1	7	8	(7,8]	0	0	1	...	1

Note that this data augmentation substantially increases the data set size: in this case, the original data from Table 4 contain $n = 131$ individuals ($=$rows), while the data in piece-wise exponential format have 5 392 rows, if all unique event and censoring times are used as interval cut-points. In addition, several auxiliary variables are constructed, see Table 5. Based on the data from Table 5 and using the `glm` function, a simple baseline PEM can be fitted. The piece-wise constant hazards for each interval are obtained simply by treating the intervals themselves as a factor variable, which results in the model specification below (using standard dummy coding, with $(\kappa_0, \kappa_1]$ as the reference category): $\lambda(t) = \lambda_0(t) = \exp\left(\beta_0 + \sum_{j=2}^{J} \beta_{0j} I(t \in (\kappa_{j-1}, \kappa_j])\right)$, where $I(t \in (a, b])$ is the indicator function for t (taking value 1, if its argument is true, and 0 otherwise). Here, β_0 represents the log baseline hazard in interval $j = 1$ and β_{0j} the log-hazard deviation in interval j compared to interval $j = 1$. When the partition of the follow-up period uses all event and censoring times as interval cut-points (and no ties are present), the regression effects of the Cox model and the PEM are equivalent (see also Section 2.1).

Alternatively, the baseline hazard can be modelled as a regression spline using 'discretized time' t_j as the covariate. A comparison of the different cumulative baseline hazard functions is presented in Figure 2. While the cumulative hazards of the Cox PH model and the PEM are equivalent (at the interval end-points), the cumulative hazard obtained by fitting a PAMM is, not surprisingly, slightly different. Here, as in the following, the point-wise confidence intervals always refer to the PAMM estimates.

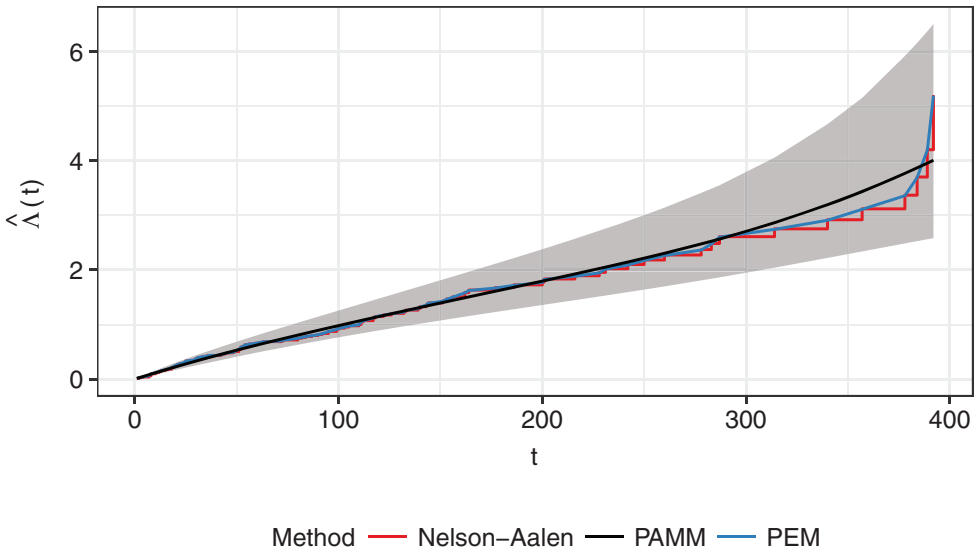


Figure 2 Comparison of the cumulative hazard estimates obtained by a Cox PH model (Nelson–Aalen), a PEM and a PAMM

3.2 Models including covariates

In this section, we incorporate conventional covariate effects into the models. We restrict ourselves to the case of time-constant covariates in this section, while the usage of TDCs is described in Section 3.4. We begin by fitting a classical Cox PH model to the original data from Table 4 using the `coxph` function, with a linear time-constant effect for the treatment variable (*trt*) and a nonlinear time-constant effect for *karno*. The corresponding hazard is then specified by $\lambda(t|\mathbf{x}) = \lambda_0(t)\exp\left(\beta_1 I(trt = 1) + f(karno)\right)$, where the *age* effect is estimated using P-splines (Eilers and Marx, 1996) and the 'optimal' degrees of freedom have been determined by Akaike's Information Criterion (AIC) (Hurvich et al., 1998), which can be done by setting the `df` argument of the `coxph` function equal to zero.

Similarly, based on data in PED format (cf. Table 5), we can fit the corresponding PAMM, simply by including a nonlinear effect $f(karno)$ into the linear predictor: $\lambda(t|x) = \exp(\lambda_0(t_j) + \beta_1 I(trt = 1) + f(karno))$. Actual estimation can be done by, for example, using the `gam` function from `mgcv`. By default, thin plate regression splines are used for such effects in `mgcv`, but here we use P-splines for direct comparison with the Cox routine. The 'optimal' effective degrees of freedom (*edf*) can be chosen by maximizing a generalized cross validation, REML or ML criterion (Wood, 2011, 2017).

Figure 3 displays the comparison of the smooth effect estimates obtained by the two approaches. Both effects exhibit a decreasing effect with increasing Karnofsky score (KS; *edf*: 4.03 for Cox; 4.02 for PAMM) and have very similar shapes. The effect

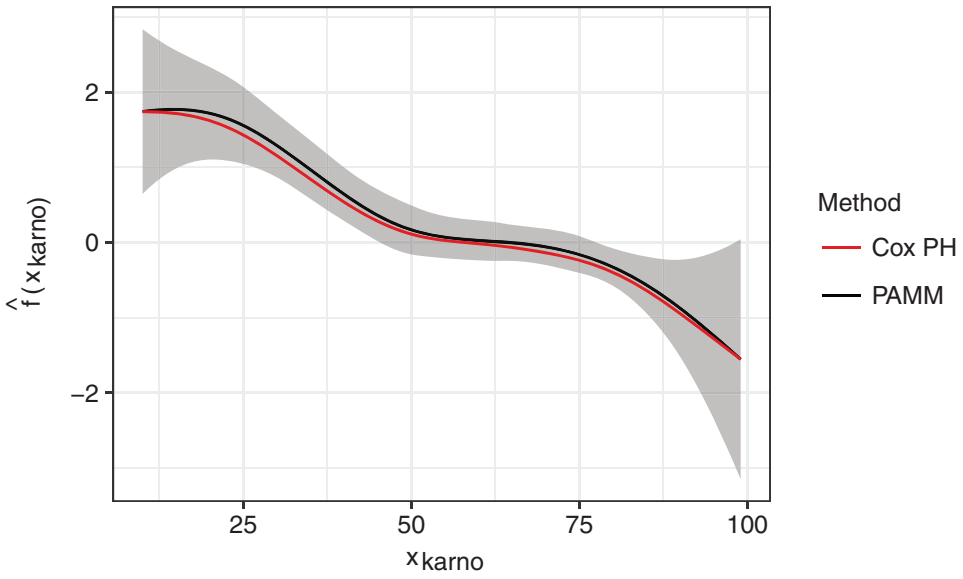


Figure 3 Comparison of nonlinear Cox PH model and PAMM estimates

of treatment is not significantly different from zero for both approaches (p-values of corresponding t-tests: 0.11 for Cox; 0.16 for PAMM).

3.3 Time-varying effects

In the following sections, we focus on time-varying effects, that is, effects of the form $f(x, t)$ for varying degrees of complexity as summarized in Table 3, and illustrate the specification of these different effect types as well as their estimation with PAMMs.

3.3.1 Stratified baseline

In many cases, it is unreasonable to assume that the baseline hazard is the same for subjects on different levels of a categorical variable. In that case, the PH assumption is violated. To avoid this issue, the so-called *stratified* PH models (Klein and Moeschberger, 1997, Ch. 9.3; also stratified Cox model in the context of Cox models) have been proposed, where a separate baseline is estimated for each sub group, while the effects of other covariates are identical for all subjects. Let z be a categorical variable with categories $k = 1, \ldots, K$. A PH model stratified with respect to z is then defined by

$$\lambda(t|z, x) = \lambda_{0k}(t) \exp(\mathbf{x}'\boldsymbol{\beta}) \,, \tag{3.1}$$

where $\lambda_{0k}(t)$ are the group-specific baseline hazards. In the Cox framework, $\boldsymbol{\beta}$ is estimated through the partial likelihood approach as usual and the baseline $\lambda_{0k}(t)$ is estimated non-parametrically for each group k.

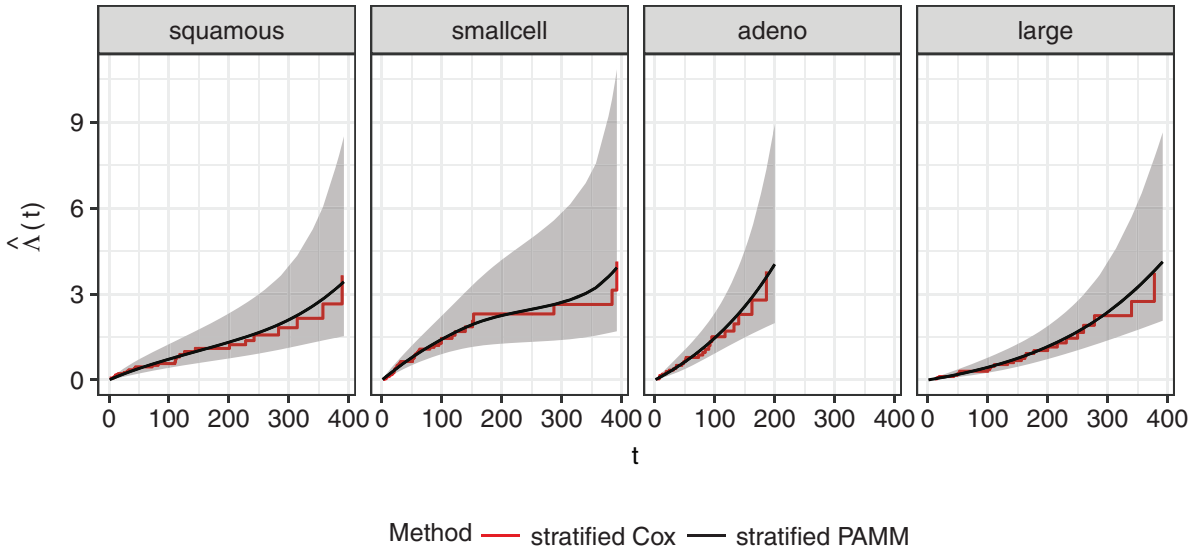


Figure 4 Cumulative hazard estimates using a stratified Cox compared to a stratified PAMM approach

In the context of PAMMs, the group-specific baseline hazards can be regarded as an interaction of the form $f(t_j) \cdot z$ between a (nonlinear) function of (discretized) time and a categorical variable, as defined in Equation (3.2):

$$\lambda(t|z, \mathbf{x}) = \exp\left(f_{0k}(t_j)I(z = k) + \mathbf{x}'\boldsymbol{\beta}\right). \tag{3.2}$$

Here, the individual baseline hazards are again represented via linear combinations of known basis functions $B(\cdot)$ and estimated basis coefficients γ, such that $f_{0k}(t_j) = \sum_{m=1}^{M} \gamma_{0km} B_m(t_j)$. In the aforementioned notation, we absorbed group-specific intercepts β_{0k} (i.e., main effects of z) into the baseline terms f_{0k} for brevity. When specifying the model in **R**, however, the group variable z must usually be included as a separate effect as well (cf. the strata vignette in Table 7).

For illustration consider patients with different cell types in the Veteran's data. Figure 4 shows the cumulative baseline hazard estimates for the different cell types using stratified Cox and stratified PAMM procedures, indicating that the two models are in good agreement. One difference between the two models, as fitted in the example, is the choice of cut-points. For the stratified Cox model, cut-points occur at respective event times in the different groups, while in the stratified PAMM model, identical cut-points are used to estimate the baseline hazards in all groups.

3.3.2 Linear, nonlinearly time-varying effects

Next, we showcase effects of the form $f(t) \cdot x$, where x is continuous. In the GAMM literature, these models are known as varying coefficient models (Hastie and Tibshirani, 1993); here the effect of x varies over time and thus constitutes a

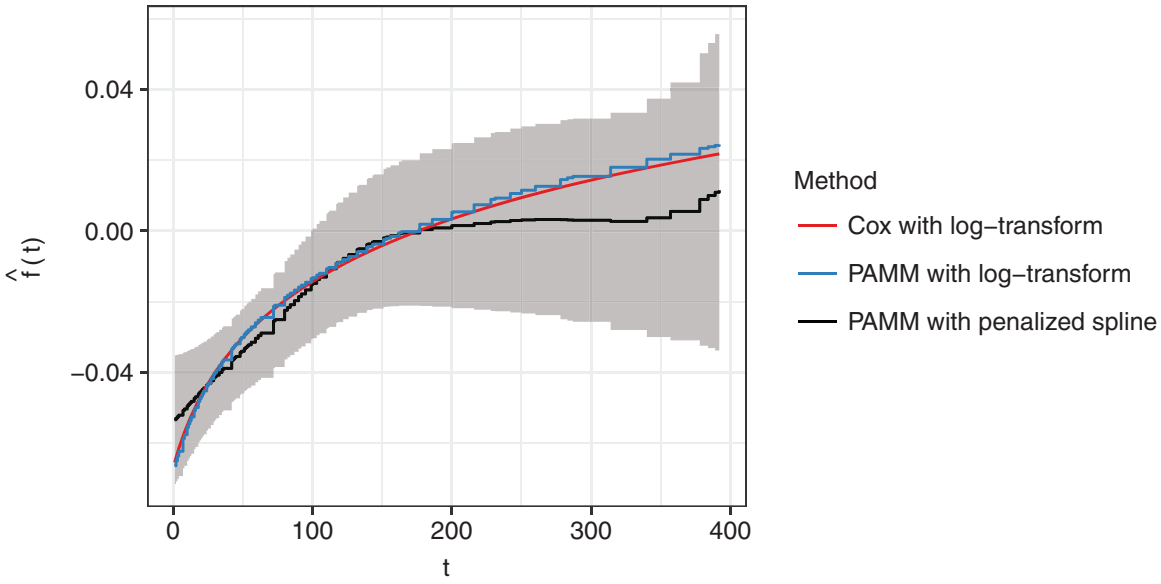


Figure 5 Comparison of estimates for the nonlinearly time-varying effect of the Karnofsky score

classical use case of time-varying effects. Note that the inclusion of time-varying effects invalidates the PH assumption, thus this additional flexibility can complicate interpretation. For illustration, we fit a time-varying effect of the KS (x_{karno}). For this association, a specific logarithmic functional form of the time variation $f(x_{karno}, t) = f(t) \cdot x_{karno} = \left(\beta_{karno} + \beta_{karno,t} \log(t + 20)\right) \cdot x_{karno}$ is frequently assumed. (Compare discussion in the `timedep` vignette of the `survival` package.)

In the PAMM context, such effects can be specified by a linear interaction between x_{karno} and a $\log(t_j + 20)$ transformation of (discretized) time. However, usually we do not know the precise functional form of $f(t_j)$ in advance and want to estimate it from the data, such that $\lambda(t|\mathbf{x}) = \exp(f_0(t_j) + f(t_j) \cdot x_{karno})$. This can be done, for example, by representing $f(t_j) \cdot x_{karno} = \sum_{m=1}^{M} \left(\gamma_m B_m(t_j)\right) \cdot x_{karno}$ in a basis function expansion as described in Section 2.3. Figure 5 shows that, in this case, the differences between the estimates using a pre-specified transformation of time and estimates of $f(t_j)$ using semiparametric regression are negligible, although the semiparametric estimate seems to 'level off' after $t = 150$.

3.3.3 Nonlinear, nonlinearly time-varying effects

In this section, we consider effects of the form $f(x, t)$, where we assume that x potentially has a nonlinear effect that also varies over time nonlinearly, estimated by penalized splines. We continue the example from Section 3.3.2, except that now $f(x_{karno}, t_j)$ will be modelled as a two-dimensional smooth function using a tensor product representation, such that $f(x_{karno}, t_j) = \sum_{m=1}^{M} \sum_{\ell=1}^{L} \gamma_{m\ell} B_m(x_{karno}) B_\ell(t_j)$.

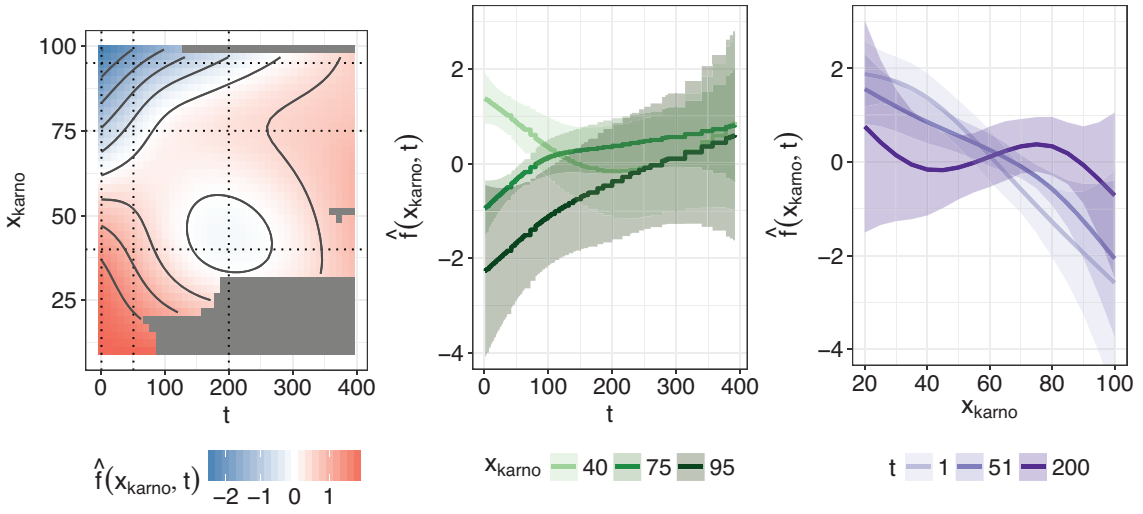


Figure 6 A heat map visualization of the nonlinear, nonlinearly time-varying effect of the KS variable on the log-hazard scale (left panel). The middle and right panel of the figure depict slices through the left panel, fixing either specific values of the KS (mid panel) or specific values of time (right panel). Grey regions in the left panel reflect combinations of x_{karno} and t where no data was observed. Intervals are point-wise ± 2 standard deviations. Note that the heat map represents a step function over time; we show a smooth surface instead for better presentation

The resulting estimate is depicted in Figure 6. Focusing on *vertical* 'slices' through the heat map on the left, we can see the effect of the KS at different times. For example, at the beginning of the follow-up ($t = 1$), low KS values are associated with higher hazards, while high KS values are associated with lower hazards, that is, very similar to the smooth, time-constant effect in Figure 5. For later time points (e.g., $t = 200$), the effect of KS is more homogeneous and generally closer to zero, even for extreme values of KS. Focusing on *horizontal* 'slices' through the heat map on the left, on the other hand, we can see how the hazard associated with different KSs changes over time. A KS of 40, for example, is associated with a much higher hazard (and, hence, decreased survival probability) at the beginning, decreases after some time and then increases again slightly, while patients with a KS close to 100 seem to have a lower hazard at the beginning, which increases towards the end of the follow-up.

This is consistent with a frequent observation in medical studies showing that the association between health scores measured at baseline and hazard rates becomes weaker or even diminishes completely over the course of the follow-up. Such a conclusion can be drawn here as well, considering the high uncertainty of the estimates for later time points. Thus, such bivariate effect functions can provide a more complete and realistic picture of time variation in the association between the hazard and the KS at baseline (increasing towards 0 for higher KS values, decreasing towards 0 for lower KS values), compared to the analyses presented in Sections 3.2 and 3.3.2.

3.4 Time-dependent covariates

Here, we describe how one can incorporate TDCs in the framework of PAMMs. Because the Veteran's data does not contain TDCs, we switch to a different standard example known from the literature, the well-known recidivism data first presented in Rossi et al. (1980). The data contains information on 432 convicts released from prison. The outcome of the study was the number of weeks until the first rearrest. Maximal follow-up time was 52 weeks. Baseline covariates include financial aid (*fin*; yes/no), *age* at time of release (in years), *race* (black/other), work experience prior to incarceration (*wexp*; yes/no), marital status (*mar;* married/unmarried), released on parol (*paro*; yes/no) and number of prior convictions (*prio*).

Furthermore, the data set contains a TDC, which indicates the weekly employment status for each subject until end of follow-up or first rearrest. For comparison, we largely follow the extensive analysis of the data presented in Fox and Weisberg (2011), who use extended Cox regression, except that we model the effects of age and prior convictions using P-splines (Eilers, 1998). In this example, the data transformation required to fit the Cox model is equivalent to the data transformation needed to apply PAMMs. More precisely, we create a data set where each subject has one row for each week of follow-up, as the employment status can potentially change every week.

A subset of the transformed data is presented in Table 6, exemplary for subjects 1 and 2. Note that the event indicator (*arrest*) is 0 up to the week of the event. Covariates other than employment remain constant over time. Note that the employment variable is 'lagged' by one week as it is unknown if unemployment preceded arrest or vice versa in any given week. Consequently, the data start with the interval $(1, 2]$, as the 'lagged' employment status for week 1 (interval $(0, 1]$) would not be defined. Note that we can exclude the offset o_{ij} for these weekly data, as it is $\log(1) = 0$ for all observations.

Figure 7 displays the comparison of the extended Cox regression and PAMM applied to the recidivism data. Both the fixed coefficients (left panel) and the smooth estimates (right panel) are generally in good agreement.

However, the nonlinear estimates of the Cox model are more 'wiggly' compared to the PAMM estimates (using default settings for both algorithms), which reflects the different approaches for the selection of the optimal smoothness penalty parameters. Being employed in the previous week is clearly associated with a decreased hazard of rearrest compared to being unemployed in the previous week. So convicts that recently had a job seem to be less likely to be arrested. For both the PAMM and the Cox PH approach the remaining fixed effects are not significant. Finally, it turns out that for both modelling approaches, increased age is generally associated with a decreased hazard, whereas convicts with a high number of prior convictions are more likely be rearrested. However, while for the Cox PH approach both effects are rather wiggly, the PAMM estimates are almost linear and very smooth and, hence, somewhat easier to interpret.

Note that we only consider the employment status of the previous week, thus the marginal hazard increase in week 21 for a subject that was employed for 19 weeks and is unemployed in week 20 is the same as for a subject that was unemployed from

Table 6 Exemplary data in counting-process format needed to fit extended Cox regression and PAMMs. Subject 1 was unemployed throughout the follow-up and rearrested in week 20, subject 2 was unemployed until week 8, found employment for weeks 9 through 13 and became unemployed again thereafter (data in the table with lag = 1). In week 17, the subject was rearrested. The other covariates are only measured at baseline and remain constant

subject	start	stop	arrest	employed (lag = 1)	fin	age	⋯	mar
1	1	2	0	0	0	27	⋯	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮
1	19	20	1	0	0	27	⋯	0
2	1	2	0	0	0	18	⋯	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮
2	8	9	0	0	0	18	⋯	0
2	9	10	0	1	0	18	⋯	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮
2	13	14	0	1	0	18	⋯	0
2	14	15	0	0	0	18	⋯	0
2	15	16	0	0	0	18	⋯	0
2	16	17	1	0	0	18	⋯	0

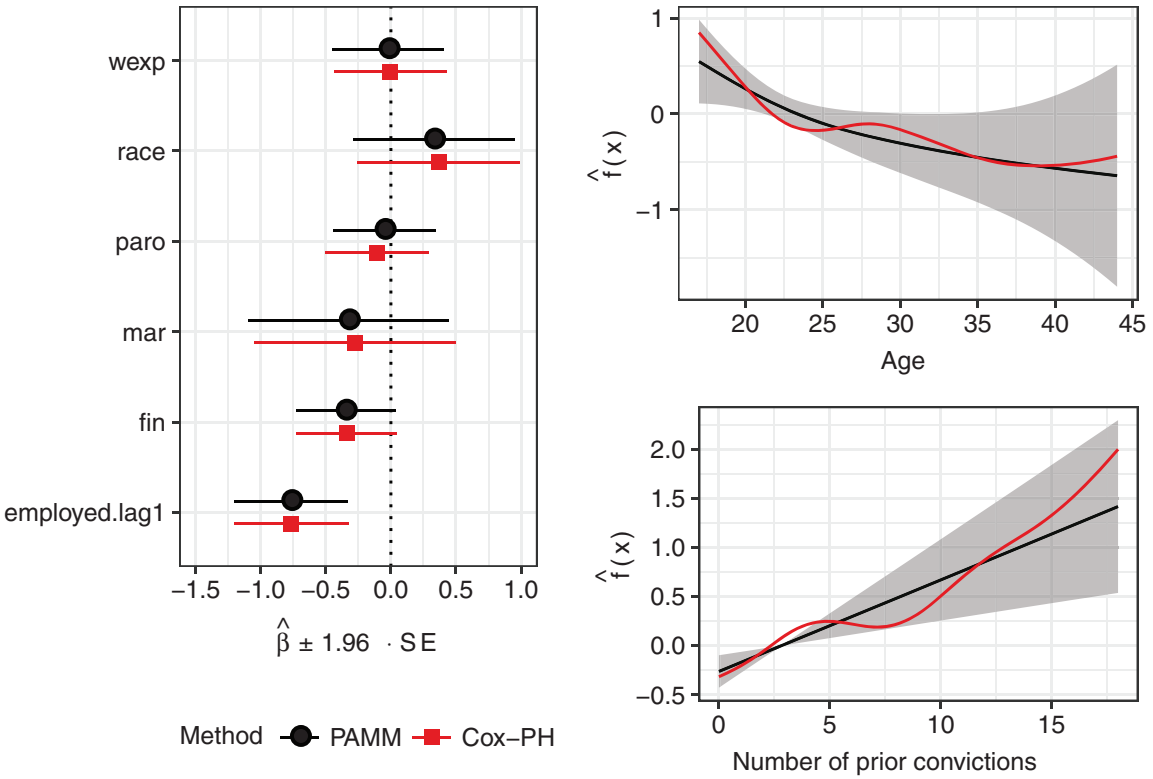


Figure 7 Left panel: Coefficient estimates using extended Cox regression and PAMM, respectively. Points indicate the estimates on the log-hazard scale. Error bars indicate the 95% confidence intervals. Right panel: Smooth estimates for age (top) and number of prior convictions (bottom)

Table 7 Overview of vignettes and their content provided in the package `pammtools`. To access the vignettes visit https://adibender.github.io/pammtools/articles/

Vignette	Description
Data transformation	Details (and respective **R**-code) on transformation of conventional time-to-event data into PED format (cf. Section 2.2)
Basics	Basic modelling and equivalence to Cox PH model
Baseline	Baseline estimation and visualization (cf. Section 2.3.1)
Splines	Linear, time-constant effects of the form $f_k(x_{i,k})$ (cf. Section 3.2)
Strata	Stratified PH models $f_k(t)I(z = k)$ (cf. Section 3.3.1)
TV effects	Details on fitting models with time-varying effects of different complexity, for example, $x_{i,k} \cdot f_k(t)$ or $f_k(x_{i,k}, t)$ (cf. Section 3.3.2 and 3.3.3)
TD covariates	Details on fitting models with time-varying covariates (cf. Section 3.4)
Frailty	Example of fitting Gaussian frailty models and comparison to the `coxme` package.
Convenience	Demonstration of convenience functions for pre-/post-processing of PEMs/PAMMs, visualization and model comparisons

the beginning. A more complex analysis would account for the complete employment history, that is, use a cumulative effect of employment, for example via the weighted cumulative exposure (WCE) approach (Sylvestre and Abrahamowicz, 2009), such that $f(x, t) = \sum_{u \leq t} = w(t - u)x(u)$, where weights $w(t - u)$ can again be estimated using penalized splines or similar approaches. An implementation of this method can be found in the corresponding **R**-package `WCE`. Bender et al. (2016) describe an implementation of such effects for PAMMs.

4 Software

While PAMMs are essentially GAMMs, and any statistical software (or programming language) could be used to fit these models, the technicalities of data transformation, interpretation and visualization often hinder their application by practitioners in our experience. Therefore, in addition to this tutorial, we also provide an **R** add-on package called `pammtools` (Bender and Scheipl, 2017); for the most current version of the package visit https://adibender.github.io/pammtools/ which provides convenience functions that facilitate the application of PAMMs. Individual sections and concepts described in this article are accompanied by vignettes on the `pammtools` homepage that illustrate the concrete application in the statistical programming language **R** (R Core Team, 2016). Table 7 gives an overview of all vignettes currently available on the site. Most conveniently, the vignettes can be accessed directly from the homepage at https://adibender.github.io/pammtools/articles. Additionally, you can call `?pammtools` to obtain a package overview and links to the individual vignettes. The package (and its vignettes) thus also serves as the code supplement for this article (see also Bender et. al 2018).

Although this article is focused on **R** and on fitting the PEMs/PAMMs with the `mgcv` package, any software that can fit GL(M)Ms or GA(M)Ms could be used to fit

this model class. For example, the standard `glm` function could be used to fit PEMs, but does not offer many of the advanced functions from `mgcv::gam` and offers no penalized estimation of smooth effects. Additionally, the package `pch` provides the function `pchreg`, that fits a variation of PEMs for right-censored and left-truncated data using a custom routine, but does not offer penalized estimation or other convenience functions for pre- and post-processing (Frumento, 2016). If random effects (or frailty terms, as they are usually referred to in survival analysis) need to be included, the `lme4` library (Bates et al., 2015) offers many options, although simple random effect structures are also supported by `mgcv::gam`. Note that Cox PH models are also supported by `mgcv(family="cox.ph")`, thus penalized estimation of smooth, nonlinear effects of time-constant covariates $f_k(x_k)$ are also directly available for the Cox model. However, these cannot be used to fit the extended Cox model with time-varying effects or TDCs. Memory efficient big data estimation with the function `bam` is also not available for `family="cox.ph"`.

In addition, regularization techniques such as the LASSO and model-based boosting, for example, via `glmnet` (Friedman et al., 2010) or `mboost` (Hothorn et al., 2016), respectively, can also be applied to PEMs and PAMMs. Although the `glmnet` package has recently implemented Cox models (see Simon et al., 2011), only rather simple Cox models can be fitted. For example, neither TDCs nor nonlinear effects are available. If `glmnet` is used to fit a PEM, however, TDCs can be included.

Within the framework of `mboost`, the `glmboost` routine with argument `family = CoxPH()` can be used for time-to-event data, if the PH assumption holds but does not allow for TDCs. With model-based boosting of PAMMs via `mboost`'s `gamboost()` function for fitting GAMs (compare, e.g., Hothorn and Bühlmann, 2006), the entire flexibility of boosted GAMMs is available for survival analysis, however. For more general details on boosting methods, compare also the tutorial on boosting approaches by Mayr and Hofner (2018) in this special issue.

5 Discussion

The focus of this tutorial article has been on the application of semiparametric regression in the context of continuous time survival analysis using piece-wise exponential (additive mixed) models, which we denote by PEMs and PAMMs. We first introduced the general idea of the classical PEM and then described its semiparametric extension, the PAMM. We illustrated the use of both approaches, starting with the required pre-processing and data augmentation steps and simple standard models, and then turned to successively more advanced applications. In this way, we hope to provide practitioners with a useful addition to their methodological toolbox for time-to-event data analysis, which is firmly embedded in the familiar context of GAMMs, allowing them to exploit the robust, well-documented and highly developed software implementations available for this model class.

The results of PEMs and PAMMs are generally in good agreement with conventional Cox-type modelling for most applications, and the concrete choice of model type can depend on many factors, including familiarity, availability of software implementations, data structure and so on. In the following, we summarize some of the strengths of PAMMs that may guide the decision process:

- Firmly grounded in modern penalization methods, PAMMs are a legitimate method for the analysis of time-to-event data and not just a 'hack' that can be used in case other options fail. For example, in Section 3, we demonstrated that we can use PEMs/PAMMs to obtain estimates very similar to the estimates of Cox-type models. While it is true that for simple use cases where the PH assumption holds the Cox model has a clear advantage over PAMMs in terms of computational cost since the data does not have to be expanded, these advantages disappear whenever time-varying effects or TDCs need to be included (cf. Sections 3.3 and 3.4). In our experience, the differences in computation time for standard applications are usually negligible anyway.
- While Cox models may have computational advantages in simpler use cases (e.g., when the PH assumption holds and the data do not have to be expanded), their computational advantage diminishes whenever time-varying effects or TDCs need to be included (cf. Sections 3.3 and 3.4). In fact, PAMMs could be more efficient in some cases, as the extended Cox model requires splits at each event time, while split points for PAMMs could be chosen more crudely.
- In the past, methodological and algorithmic advances in semiparametric regression first became available in the framework of GAMMs. Thus, for example, the important innovations implemented in the `mgcv` package, such as REML-optimal estimation of penalized effects (Wood et al., 2016), reliable and generally applicable tests for semiparametric terms (Wood, 2012), locally adaptive spline smooths (Wood, 2011), double-penalty variable selection strategies that can shrink effect estimates to zero (Marra and Wood, 2011) similar to the (group) LASSO (Meier et al., 2008), highly memory efficient estimation on huge data sets (Wood et al., 2016), and many more, are directly available for the estimation of PAMMs. Similarly, in the past, regularization approaches such as boosting became available for GAMMs long before being ported to Cox-type models.
- In Section 3.3.2, it was shown how easily linear but nonlinearly time-varying effects of the form $f(t) \cdot x$ can be incorporated in the framework of PAMMs, being directly available within the syntax of `gam`. In contrast, in conventional Cox modelling either a particular functional form of the time variation has to be assumed (as illustrated in Section 3.3.2) or further pre-processing steps become necessary, such as manual construction of the corresponding spline design matrices in combination with suitable penalization (see, e.g., the simulation study in Groll et al., 2017, for an application of this strategy).
- TDCs can be embedded in this framework naturally (cf. Section 3.4), and even more complicated effects of TDCs, for example, the WCE approach by

Sylvestre and Abrahamowicz (2009) or the DLNM of Gasparrini et al. (2017), where cumulative time-varying effects of TDCs are considered, can be directly incorporated with fairly little additional effort.

With respect to assessment of model quality in the context of PEMs/PAMMs, we advise not to use conventional measures like, for example, the deviance, but rather use dedicated measures developed specifically for survival analysis such as the Brier score or the concordance index (Gerds et al., 2013), which have the secondary advantage of being directly comparable with evaluation metrics for other model classes for survival analysis. Finally, note that in this article we did not address any issue regarding estimation or inference, for which we refer to Wood (2017, 2011).

References

Andersen PK, Borgan O, Gill R and Keiding N (1992) *Statistical Models Based on Counting Processes*. Berlin and New York, NY: Springer-Verlag.

Argyropoulos C and Unruh ML (2015) Analysis of time to event outcomes in randomized controlled trials by generalized additive models. *PLoS ONE*, **10**, e0123784. doi:10.1371/journal.pone.0123784

Bates D, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. doi:10.18637/jss.v067.i01

Bender A and Scheipl F (2017, November 14). *adibender/pammtools: v0.0.3.2* (Version v0.0.3.2). Zenodo. URL http://doi.org/10.5281/zenodo.1048832

Bender A, Groll A and Scheipl F (2018, January 14). *adibender/pammtutorial-smj: Release v1.0.1* (Version v1.0.1). Zenodo. URL http://doi.org/10.5281/zenodo.1147058

Bender A, Scheipl F, Küchenhoff H, Day AG and Hartl W (2016) *Modeling exposure-lag-response associations with penalized piecewise exponential models* (Technical report 108). Ludwig-Maximilians-University URL https://epub.ub.uni-muenchen.de/32010/.

Berger M and Schmid M (2018) Semiparametric regression for discrete time-to-event data. *Statistical Modelling*, **18**.

Clayton DG (1983) Fitting a general family of failure-time distributions using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **32**, 102–109. doi:10.2307/2347288

Cox DR (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, **B 34**, 187–220.

Demarqui FN, Loschi RH and Colosimo EA (2008) Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis*, **14**, 333–56. doi:10.1007/s10985-008-9086-0

Eilers PHC (1998) Hazard smoothing with B-splines. *Proceedings of the 13th International Workshop on Statistical Modelling*, New Orleans, La, pages 200–07..

Eilers PHC and Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121. doi:10.1214/ss/1038425655

Fox J and Weisberg HS (2011) *An R Companion to Applied Regression*. Thousand Oaks, CA: SAGE. ISBN 978-1-4129-7514-8.

Friedman M (1982) Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, **10**, 101–13.

Friedman J, Hastie T and Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1–22.

Frumento P (2016) *pch: Piecewise constant hazards models for censored and truncated data*. R package version 1.3. URL https://CRAN.R-project.org/package=pch

Gasparrini A, Scheipl F, Armstrong B and Kenward MG (2017) A penalized framework for distributed lag non-linear models. *Biometrics*. doi:10.1111/biom.12645

Gerds TA, Kattan MW, Schumacher M and Yu C (2013) Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine*, **32**, 2173–84. doi:10.1002/sim.5681

Groll A, Hastie T and Tutz G (2017) Selection of effects in Cox frailty models by regularization methods. *Biometrics*, **73**, 846–56.

Guo G (1993) Event-history analysis for left-truncated data. *Sociological Methodology*, **23**, 217–43. doi:10.2307/271011

Hastie T and Tibshirani R (1993) Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 757–96. doi:10.2307/2345993

Holford TR (1980) The analysis of rates and of survivorship using log-linear models. *Biometrics*, **36**, 299–305. doi:10.2307/2529982

Hothorn T and Bühlmann P (2006) Model-based boosting in high dimensions. *Bioinformatics*, **22**, 2828–29.

Hothorn T, Bühlmann P, Kneib T, Schmid M and Hofner B (2016) *mboost:* Model-based boosting. R package version 2.7-0. URL http://CRAN.R-project.org/package=mboost/

Hurvich CM, Simonoff JS and Tsai C (1998) Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**, 271–93.

Kalbfleisch J and Prentice R (1980) *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley.

Klein JP and Moeschberger ML (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York, NY: Springer.

Laird N and Olivier D (1981) Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, **76**, 231–40. doi:10.2307/2287816

Marra G and Wood SN (2011) Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, **55**, 2372–87. doi:10.1016/j.csda.2011.02.004

Martinussen T and Scheike TH (2006) *Dynamic Regression Models for Survival Data*. New York, NY: Springer.

Mayr A and Hofner B (2018) Boosting for statistical modelling: A non-technical introduction. *Statistical Modelling*, **18**.

Meier L, Van de Geer S and Bühlmann P (2008) The group LASSO for logistic regression. *Journal of the Royal Statistical Society*, **B 70**, 53–71.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing URL https://www.R-project.org/

Rodríguez-Girondo M, Kneib T, Cadarso-Suárez C and Abu-Assi E (2013) Model building in nonproportional hazard regression. *Statistics in Medicine*, **32**, 5301–14. doi:10.1002/sim.5961

Rossi PH, Berk RA and Lenihan KJ (1980) *Money, Work, and Crime: Experimental Evidence*. New York: Academic Press.

Ruppert D, Wand MP and Carroll RJ (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.

Sennhenn-Reulen H and Kneib T (2016) Structured fusion LASSO penalized multi-state models. *Statistics in Medicine*. doi:10.1002/sim.7017

Simon N, Friedman J, Hastie T and Tibshirani R (2011) Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, **39**, 1.

Sylvestre M-P and Abrahamowicz M (2009) Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine*, **28**, 3437–53. doi:10.1002/sim.3701

Therneau TM (2015) *A package for survival analysis in S*. R package version 2.38. URL http://cran.us.r-project.org/web/packages/survival/index.html

Thomas L and Reyes EM (2014) Tutorial: Survival estimation for Cox regression models with time-varying coefficients using

SAS and R. *Journal of Statistical Software, Code Snippets*, **61**. URL http://www.jstatsoft.org/v61/c01

Whitehead J (1980) Fitting Cox's regression model to survival data using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **29**, 268–75. doi:10.2307/2346901

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**, 3–36.

——— (2012) On p-values for smooth components of an extended generalized additive model. *Biometrika*, **100**, 221–28. doi:10.1093/biomet/ass048

——— (2017) *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*. URL http://cran.r-project.org/web/packages/mgcv/index.html

Wood SN, Li Z, Shaddick G and Augustin NH (2016) Generalized additive models for gigadata: Modelling the UK black smoke network daily data. *Journal of the American Statistical Association*, 1–40. doi:10.1080/01621459.2016.1195744

Wood SN, Pya N and Saefken B (2016) Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, **111**, 1548–63. doi:10.1080/01621459.2016.1180986

Part II

Exposure-lag-response Associations

Chapter 3

Penalized Estimation of complex exposure-lag-response associations

Chapter 3 introduces a general framework for the representation and estimation of exposure-lag-response associations. The methods are used to investigate the association between caloric intake and short term survival in critically ill patients.

Contributing article:

Bender, A., Scheipl, F., Hartl, W. H., Day, A. G., and Küchenhoff, H. (2018b). Penalized estimation of complex, non-linear exposure-lag-response associations. *Biostatistics*.

Copyright:

Author contributions:

Andreas Bender was responsible for the methodological development of the general framework for cumulative effects in time-to-event data analysis. Wolfgang W. Hartl, Fabian Scheipl, Andreas Bender and Helmut Küchenhoff developed the statistical analysis strategy for the application example. Andreas Bender and Fabian Scheipl performed the statistical modeling and inference. Andrew G. Day provided the data base and valuable input. Andreas Bender performed the simulation studies and prepared the manuscript. All authors critically revised and contributed to the manuscript.

Supplementary material available at:

`http://biostatistics.oxfordjournals.org`
`https://zenodo.org/record/1203238`

Biostatistics (2018) **00**, 00, *pp.* 1–17
doi:10.1093/biostatistics/kxy003

Penalized estimation of complex, non-linear exposure-lag-response associations

ANDREAS BENDER*

Statistical Consulting Unit, StaBLab, Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 Munich, Germany

andreas.bender@stat.uni-muenchen.de

FABIAN SCHEIPL

Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 Munich, Germany

WOLFGANG HARTL

Department of General, Visceral, Transplantation, and Vascular Surgery, University School of Medicine, LMU Munich, Grosshadern Campus, Marchioninistraße 15, 81377 Munich, Germany

ANDREW G. DAY

Clinical Evaluation Research Unit, Kingston General Hospital, KGH Research Institute, 76 Stuart Street, Kingston, Ontario K7L 2V7, Canada

HELMUT KÜCHENHOFF

Statistical Consulting Unit, StaBLab, Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 Munich, Germany

SUMMARY

We propose a novel approach for the flexible modeling of complex exposure-lag-response associations in time-to-event data, where multiple past exposures within a defined time window are cumulatively associated with the hazard. Our method allows for the estimation of a wide variety of effects, including potentially smooth and smoothly time-varying effects as well as cumulative effects with leads and lags, taking advantage of the inference methods that have recently been developed for generalized additive mixed models. We apply our method to data from a large observational study of intensive care patients in order to analyze the association of both the timing and the amount of artificial nutrition with the short term survival of critically ill patients. We evaluate the properties of the proposed method by performing extensive simulation studies and provide a systematic comparison with related approaches.

Keywords: Cumulative effects; Exposure-lag-response association; Penalized additive models; Reproducible research; Survival analysis; Time-dependent covariates.

*To whom correspondence should be addressed.

1. INTRODUCTION

Patients admitted to an intensive care unit (ICU) often require mechanical ventilation (MV) and artificial nutrition (enteral or parenteral). The optimal timing and amount of artificial nutrition, however, is unclear and previous observational and randomized studies have not yielded consistent results (for details, see Tables S5 and S6 in Appendix A.9 of supplementary material available at *Biostatistics* online). Modeling the association between caloric intake and survival is particularly challenging for the following reasons: First, the amount of artificial caloric intake during ICU stay varies daily on a per patient basis. Second, effects of artificial nutrition may vary over time, and hazard rates at a particular point in time may depend on multiple past caloric intakes (i.e., the exposure history). Third, the caloric intake may have a delayed impact on the outcome, and may "wear off" after some time. Consequently, contradictory results of recent observational studies may be explained by their inability to account for the interplay between the timing and amount of artificial nutrition. Older observational studies have only considered the average caloric intake over a defined period (usually 1–2 weeks) and ignored the exact timing of caloric intake. In general, our proposal aims at modeling the cumulative effect of past exposure as a function of both the timing and amount of exposure which might additionally vary over the follow-up time. In the context of the application, this time-variation means that the contribution to the hazard of a given amount of nutrition on Day 4 to the hazard at Day 7 may be different than that of an identical amount of nutrition on Day 6 to the hazard at Day 9, even though the time between exposure and risk is the same in both cases.

Let us define some necessary terminology. We denote effects of time-constant covariates that can vary over time as *time-varying effects*. Variables whose values change over time (here nutrition on the ICU), are referred to as *time-dependent covariates* (TDC). We denote the follow-up time by t and the times at which the TDC was observed by t_e. The value of the TDC at time t_e is then denoted by $z(t_e)$ and a subject's exposure history by $\mathbf{z} = (z(t_{e_1}), z(t_{e_2}), \ldots)$. The *lag* time t_{lag} is defined as the length of the delay until the TDC recorded at time t_e starts to affect the hazard. The lead time t_{lead} defines the duration of the effect of the TDC recorded at t_e. That is, $z(t_e)$ can affect the hazard for all $t \in [t_e + t_{\text{lag}}, t_e + t_{\text{lag}} + t_{\text{lead}}]$ and, vice versa, the hazard at time t is affected by exposures $\{z(t_e) : t_e \in [t - t_{\text{lag}} - t_{\text{lead}}, t - t_{\text{lag}}]\}$. In this work, we propose a novel approach for the modeling of such *exposure-lag-response associations* (ELRAs; Gasparrini, 2014) that extends previous research by (i) taking into account all three dimensions relevant to the effect of a time-dependent exposure: the timing of exposures t_e as well as their amounts recorded in a subject's exposure history $\mathbf{z}$, whose effect can vary over follow-up time t and (ii) by allowing for more flexible definitions of the window of effectiveness defined by t_{lag} and t_{lead}. Thus, we present a general framework for the estimation of effects like

$$g(\mathbf{z}, t) = \int_{t - t_{\text{lag}}(t_e) - t_{\text{lead}}(t_e)}^{t - t_{\text{lag}}(t_e)} h(t, t_e, z(t_e)) \mathrm{d}t_e, \tag{1.1}$$

where $g(\mathbf{z}, t)$ denotes the cumulative effect of exposure history $\mathbf{z}$ on the log-hazard at time t. The term $h(t, t_e, z(t_e))$ denotes individual contributions of exposures $z(t_e)$ recorded at specific exposure times t_e. We refer to these contributions as *partial effects*. The cumulative effect $g(\mathbf{z}, t)$ is defined as the integral of these partial effects within a pre-specified time window defined by lag and lead times and can be approximated by a weighted sum over the respective partial effects.

Previous approaches are special cases of this general definition: *weighted cumulative exposure* (WCE) models (Sylvestre and Abrahamowicz, 2009) assume that $h(t, t_e, z(t_e)) \overset{!}{=} \tilde{h}(t - t_e) z(t_e) \; \forall \, t, \; t_e$, i.e., partial effects that are linear in $z(t_e)$ and depend only on the latency $t - t_e$. Sylvestre and Abrahamowicz (2009) estimate $\tilde{h}(t - t_e)$ using B-Spline basis expansions and smoothness is controlled with BIC-based knot selection. Berhane *and others* (2008) use unpenalized B-Spline tensor product bases to model the relative risk of dying as a bivariate non-linear function of radon exposure $z(t_e)$ and latency $t - t_e$, i.e., they set

$h(t, t_e, z(t_e)) \overset{!}{=} \tilde{h}(t - t_e, z(t_e)) \, \forall \, t, \; t_e$. Estimation is performed using logistic regression models, which only applies to discrete-time models (or discretized time) and they do not provide uncertainty estimates for the bivariate surface estimate. Gasparrini (2014) defines the framework of distributed lag non-linear models (DLNM), where partial effects are specified as in Berhane *and others* (2008) and a penalized likelihood estimation approach for DLNMs is described in Gasparrini *and others* (2017). This DLNM can be obtained as a special case of (1.1) in the same way as for Berhane *and others* (2008). In all previous approaches, the weight of exposure $z(t_e)$ only depends on the latency $t - t_e$, not the specific combination of t and t_e. Thus, for example, $h(t = 30, t_e = 3) \overset{!}{=} h(t = 40, t_e = 13) \overset{!}{=} \tilde{h}(t - t_e = 27)$, which can be a perfectly reasonable assumption in many cases, but is often unknown a priori. Relaxing this assumption leads to more flexible versions of the WCE and DLNM. A more general WCE model thus could be defined using partial effects $h(t, t_e) \cdot z(t_e)$, which can not be defined as a special case of the DLNM. In Section 3, we apply this extension of the WCE to the nutrition data with a three categorical $z(t_e)$. A more flexible version of the DLNM, defined by partial effects $h(t, t - t_e, z(t_e))$, is used in simulation study Part A, Section 4.2. When we define ELRAs in Section 2.2.4, we allow additional flexibility, as t_{lag} and t_{lead} can also be functions of exposure time t_e, thus the number of past exposures that contribute to the cumulative effect can also depend on (exposure) time.

The remainder of this article is structured as follows. Our model, which is an extension of the piece-wise exponential model (PEM), is described in detail in Section 2. The proposed method for estimating ELRAs is presented in Section 2.2.4. Section 3 shows an application of our approach to observational data of almost 10 000 critically ill patients. In Section 4, we present results of an extensive simulation study to assess properties of the proposed modeling approach and to investigate its behavior if modeling assumptions are violated. A summary and discussion are presented in Section 5. Details regarding reproducibility of our analyses can be found in Section 6.

2. Models and methods

In the following, we outline the proposed framework for fitting ELRAs, starting from simpler models and gradually increasing their complexity. This motivates the general flexibility of the proposed model class (cf. Argyropoulos and Unruh, 2015) and helps to understand ELRAs as natural extensions of time-varying effects for TDCs.

2.1. *Piece-wise exponential model*

Given a partition of the follow-up period $(0, t_{\max}]$ into J intervals with $J + 1$ cut-points $0 = \kappa_0 < \ldots < \kappa_J = t_{\max}$, where $t_{\max}$ is the maximal follow-up time, and assuming the baseline hazard rate $\lambda_0(t)$ in each interval j to be constant, such that $\lambda_0(t) = \lambda_{0j}, \forall t \in (\kappa_{j-1}, \kappa_j], t > 0$, a proportional hazards model for subjects $i = 1, \ldots, n$ can be written as

$$\log(\lambda_i(t|\mathbf{x}_i)) = \log(\lambda_{0j}) + \mathbf{x}_i'\boldsymbol{\beta} \; \forall \; t \in (\kappa_{j-1}, \kappa_j], \tag{2.1}$$

with $\mathbf{x}_i' = (x_{i,1}, \ldots, x_{i,P})$ a row-vector of P time-constant covariates.

This is the PEM of Whitehead (1980) and Friedman (1982), whose likelihood is equivalent to that of a Poisson generalized linear model (GLM)

$$\log(\mathbb{E}(y_{ij}|\mathbf{x}_i)) = \log(\lambda_{0j}) + \mathbf{x}_i'\boldsymbol{\beta} + \log(t_{ij}) \tag{2.2}$$

with (i) one observation for each interval $j = 1, \ldots, J$ under risk for subject i, (ii) responses as event indicators for subject i for interval j, i.e., $y_{ij} = I(t_i \in (\kappa_{j-1}, \kappa_j] \wedge t_i = T_i)$, and (iii) $t_{ij} = \min(t_i - \kappa_{j-1}, \kappa_j -$

κ_{j-1}), representing the time subject i spent under risk in interval j. We define $t_i := \min(T_i, C_i)$ as the observed right-censored time under risk for subject i with event time T_i and censoring time C_i, which is assumed to be non-informative for this model class.

In practice, when fitting the respective Poisson regression model, $\log(\lambda_{0j})$ is incorporated in the linear predictor $\mathbf{x}_i'\boldsymbol{\beta}$ and $\log(t_{ij})$ enters as an offset. In the following sections, our model specifications for the log-hazard rate do not contain the offset term, but the offset must be included at the estimation stage. This model structure lends itself easily to include TDCs, as a covariate can change its value at each κ_j, i.e., *additional* interval cut-points can be chosen at the time-points at which a change in the TDC is recorded (see Section 2.2.1 for more details on the choice of cut-points). Then (2.1) can be extended to $\log(\lambda_i(t|\mathbf{x}_{ij})) = \log(\lambda_{0j}) + \mathbf{x}_{ij}'\boldsymbol{\beta}$. Additionally, time-varying effects can be incorporated by creating a TDC for time itself, e.g., by using the interval midpoints $\tilde{t}_j := (\kappa_j - \kappa_{j-1})/2$, and including interaction terms of selected covariates with $\tilde{t}$, or transformations thereof, in the linear predictor.

2.2. *Piece-wise exponential additive model*

Model (2.2) is a GLM, in which covariate effects are assumed to be linear in x and constant or linear over t as well as non-cumulative, and observations are assumed to be independent. In order to remove these restrictions, we introduce the class of *p*iece-wise exponential *a*dditive *m*ixed *m*odels (PAMMs) (cf. Argyropoulos and Unruh, 2015), in analogy to the extension of GLMs to generalized additive mixed models (GAMMs). By doing so, we can simultaneously (i) achieve reliable estimates of the baseline hazard parameters $\lambda_{0j}, j = 1, \ldots, J$ even for very fine partitions $[\kappa_0, \ldots, \kappa_J]$ of the follow-up with large J (Section 2.2.1), (ii) include random frailty effects to model the heterogeneity of and dependence between observations with a known grouping structure, e.g., patients from different hospitals in a multicenter study (Section 2.2.2), (iii) include non-linearly time-varying and non-linear effects of both time-constant or TDCs (Section 2.2.3), (iv) include cumulative, time-lagged, and time-varying effects (i.e., ELRAs) of TDCs (Section 2.2.4).

The basic idea—detailed in the subsections below—is to use penalized estimation of spline basis representations of the respective effects or rates in order to obtain models that are highly flexible and yet allow for reliable inference (cf. our Section 2.3, Argyropoulos and Unruh, 2015, p. 13f). In particular, both the time variation and the non-linear functional shape of covariate effects do not have to be specified *a priori*, instead they are estimated from the data.

2.2.1. *Baseline hazard.* In Whitehead (1980)'s definition of PEMs (2.2), the baseline hazard $\lambda_0(t)$ is a step function and interval-specific hazard rates λ_{0j} are estimated by including interval-specific dummy variables in the model matrix. The choice of interval cut-points can be problematic (Demarqui *and others*, 2008) and choosing a very high number of cut-points increases the number of parameters that need to be estimated. The latter reduces the stability and precision of the individual estimates $\hat{\lambda}_{0j}$ and often leads to implausibly large changes in the estimated baseline hazard from one interval to the next. To avoid these issues, we estimate the baseline log-hazard step function with a spline evaluated at the interval midpoints $\tilde{t}_j$. First, this means that a very fine partition of the follow-up can be used since the number of parameters for the baseline hazard is given by the dimension of the spline basis, not by the number of intervals J, thus reducing the importance of the location of cut-points. Second, given a sufficiently large spline basis and a fine partition of the intervals, the baseline hazard can be estimated very flexibly, as it changes at each of the closely adjacent interval cut-points. Such closely adjacent cut-points also make the implausibility of restricting the estimated hazard rate to a *step function* with steps at κ_j negligible for practical purposes. Third, large hazard rate changes between adjacent intervals that are not strongly supported by the data are avoided since the estimate of $\log(\lambda_{0j}) = f(\tilde{t}_j)$ is suitably penalized (cf. Section 2.3).

Table 1. *Overview of possible effect specifications. $h_p(\cdot)$ denotes a smooth function of its arguments*

	Effect specification	Description
$f_p(x_p, t) =$	$\beta_p x_{i,p}$:	Linear, time-constant effect
$f_p(x_p, t) =$	$h_p(x_{i,p})$:	Smooth, time-constant effect
$f_p(x_p, t) =$	$\beta_p x_{i,p} + \beta_{p:t}(x_{i,p} \cdot t)$:	Linear, linearly time-varying effect
$f_p(x_p, t) =$	$h_p(x_{i,p}) \cdot t$:	Smooth, linearly time-varying effect
$f_p(x_p, t) =$	$x_{i,p} \cdot h_p(t)$:	Linear, smoothly time-varying effect
$f_p(x_p, t) =$	$h_p(x_{i,p}, t)$:	Smooth, smoothly time-varying effect

2.2.2. *Frailties and random cluster effects.* For data with a known grouping structure, we can extend (2.2) by Gaussian random effects (i.e., frailties) affecting the group-specific hazard rates. Defining $\ell = 1, \ldots, L$ as the index for different groups and ℓ_i as the grouping level to which subject i belongs, we write

$$\log\left(\lambda_i(t|\mathbf{x}_i, \ell_i)\right) = \log(\lambda_0(t)) + \mathbf{x}_i'\boldsymbol{\beta} + b_{\ell_i}$$

with Gaussian random effects b_{ℓ_i} that capture heterogeneity and dependence induced by the grouping structure. For example, our application contains random effects for the ICUs where the patients were treated. Their inverse variance is estimated from the data as one of the model's smoothing parameters (cf. Section 2.3).

2.2.3. *Flexible covariate effects.* The effects $\mathbf{x}_i'\boldsymbol{\beta}$ of time-constant covariates $\mathbf{x}_i$ in equation (2.1) denote simple linear and time-constant associations with the log-hazard rate. Much more generally, PAMMs can contain possibly non-linear, possibly time-varying effects of time-constant covariates on the log hazard, denoted by $f_p(x_{i,p}, t)$. We extend equation (2.1) to

$$\log\left(\lambda_i(t|\mathbf{x}_i, \ell_i)\right) = \log(\lambda_0(t)) + \sum_{p=1}^{P} f_p(x_{i,p}, t) + b_{\ell_i}.$$

Note that, like the baseline, all time-varying effects in a PAMM are step functions over the partition given by $\kappa_j, j = 0, \ldots, J$, so that $f_p(x_{i,p}, t) \equiv f_p(x_{i,p}, \tilde{t}_j)\ \forall\, t \in (\kappa_{j-1}, \kappa_j]$. Table 1 shows the variety of covariate effects subsumed into this notation, which range from the simple linear, time constant effects in (2.2) to varying coefficients $x_{i,p}h_p(t)$ or $h_p(x_{i,p})t$ (Hastie and Tibshirani, 1993) to non-linear, smoothly time-varying covariate effects $h_p(x_{i,p}, t)$ modeled as bivariate function surfaces, parameterized as tensor product smooths (Wood *and others*, 2013). The smooth functions $f_p(\cdot)$ can be represented as splines of the form $f_p(\cdot) = \sum_{m=1}^{M} \gamma_{m,p} B_{m,p}(\cdot)$, where $B_{m,p}$ are covariate specific (tensor product) basis functions and $\gamma_{m,p}$ are the associated spline coefficients estimated from the data controlling the effect's shape. Specification $f_p(x_{i,p}, t)$ is the most flexible and should be employed whenever prior information or domain specific knowledge regarding the relationship is absent and a sufficiently large number of cases is available for reliable estimates. Note that time-varying effects imply non-proportional hazards and that (non-)linearity is irrelevant for effects of categorical covariates represented by dummy variables. All this flexibility, of course, raises the question of model selection (cf. discussion in Section 5).

2.2.4. *Exposure-lag-response associations.* For the sake of notational simplicity, we present a model with only one ELRA of a single TDC of primary interest, the *exposure* $z_i(t_e)$. An extension to multiple

ELRAs, however, is straightforward, as will be illustrated in the application example (cf. Section 3). Denote a subject's *exposure history* by $\mathbf{z}_i = (z_i(t_{e,1}), \ldots, z_i(t_{e,Q_i}))$, with $t_{e,1}, \ldots, t_{e,Q_i}$ assumed to be sufficiently dense over t_e so that all relevant changes of z_i are recorded in $\mathbf{z}_i$.

Define the set of intervals in which hazard rates are potentially affected by exposure at time t_e as $\mathcal{J}(t, t_e) := \{(\kappa_{j-1}, \kappa_j] : \kappa_{j-1} > t_e + t_{\text{lag}}(t_e) \wedge \kappa_j \leq t_e + t_{\text{lag}}(t_e) + t_{\text{lead}}(t_e)\}$ where $t_{\text{lag}}(t_e)$ and $t_{\text{lead}}(t_e)$ are the lag and lead times as defined in the introduction (cf. Section 1, Equation (1.1)). Vice versa, exposure times potentially affecting hazards in interval j are given by $\mathcal{T}_e(j) := \{t_e : (\kappa_{j-1}, \kappa_j] \in \mathcal{J}(t, t_e)\}$. In the following, we refer to the set $\mathcal{T}_e(j)$ as *lag-lead window* (see top of Figure 1 for examples and intuition). Currently, lag and lead times must be defined *a priori*. In the general definition, $t_{\text{lag}}(t_e)$ and $t_{\text{lead}}(t_e)$ can depend on the exposure time, i.e., the width of the lag-lead window can change over time. In many applications, however, constant lag and lead times will be a reasonable assumption, which can be incorporated in our approach directly, by defining $t_{\text{lag}}(t_e) = t_{\text{lag}}\ \forall t_e$ and $t_{\text{lead}}(t_e) = t_{\text{lead}}\ \forall t_e$.

An ELRA $g(\mathbf{z}, t)$ represents the cumulative, time-varying effect of $\mathbf{z}$ on the log-hazard, so we define its contribution to the model's additive predictor as

$$g(\mathbf{z}_i, t) = \int_{\mathcal{T}_e(j)} h(\tilde{t}_j, t_e, z_i(t_e))\mathrm{d}t_e \approx \sum_{q:t_{e,q} \in \mathcal{T}_e(j)} \Delta_q h(\tilde{t}_j, t_{e,q}, z_i(t_{e,q})) \quad \forall\, t \in (\kappa_{j-1}, \kappa_j], \tag{2.3}$$

where $h(t, t_e, z(t_e))$ is the partial effect of exposure value $z(t_e)$ observed at t_e on the hazard in interval j to which t belongs. The total (cumulative) ELRA effect at time t is then given by the integral of all partial effects of exposures within the window of effectiveness given by $\mathcal{T}_e(j)$, approximated by a weighted sum. Like all time-varying effects in a PAMM, $g(\mathbf{z}_i, t)$ is a step function over the partition of t given by the κ_j, evaluated at the interval mid points $\tilde{t}_j, j = 1, \ldots, J$. Quadrature weights $\Delta_q = t_{e,q} - t_{e,q-1}$ for numerical integration are given by the time between two consecutive exposure measurements.

If we restrict the ELRA to be linear in the exposure, i.e., $h(z_i(t_e), t_e, t) = \tilde{h}(t_e, t) \cdot z_i(t_e)$ we can simplify (2.3) to $g(\mathbf{z}_i, t) \approx \sum_{q=1}^{Q} \tilde{\Delta}_{i,q} \tilde{h}(t_{e,q}, t)$, with $\tilde{\Delta}_{i,q} = \begin{cases} z_i(t_{e,q})\Delta_q & \text{if } t_{e,q} \in \mathcal{T}_e(j) \\ 0 & \text{else} \end{cases}$. In order to set up a spline basis for the bivariate function $\tilde{h}(t_e, t)$, we use a tensor product B-spline basis with marginal bases $B_m(t_e), m = 1, \ldots, M$ and $B_k(t), k = 1, \ldots, K$ defined over the exposure and hazard time domains, respectively. Such tensor product bases allow the construction of multivariate basis functions over disparate dimensions, where M and K delimit the maximal complexity of the ELRA over t_e and t, respectively. Thus, we write $\tilde{h}(t_e, t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \gamma_{m,k} B_m(t_e) B_k(t)$.

Combining these two equations, the linear ELRA's basis function expansion is then given by

$$g(\mathbf{z}_i, t) \approx \sum_{m=1}^{M} \sum_{k=1}^{K} \gamma_{m,k} \tilde{B}_{i,m}(t_e, t) B_k(t), \tag{2.4}$$

where $\tilde{B}_{i,m}(t_e, t) = \sum_{q=1}^{Q} \tilde{\Delta}_{i,q} B_m(t_e)$. The penalized estimation of $h(\cdot)$ (cf. Section 2.3) implies an assumption of smoothness on $\tilde{h}(t_e, t)$, which ensures that effects of exposures at consecutive exposure times t_e, t_e' are similar and that effects of a given exposure history $\mathbf{z}_i$ on the hazards in neighboring intervals j, j' are similar as well. Paired with an anisotropic penalty, the specification via a tensor product basis allows for different amounts of roughness over t_e and t. This is crucial if (i) the effect's complexity is different over the two time variables, e.g., if the timing of the exposure is largely irrelevant, but its effect is strongly time-dependent so that $\tilde{h}(t_e, t)$ is approximately constant over t_e but highly variable over t or if (ii) exposure and hazards are observed on different time domains. For example, in this article's motivating application, nutrition information is available on a daily basis only for the second to twelfth calendar days of ICU stay,

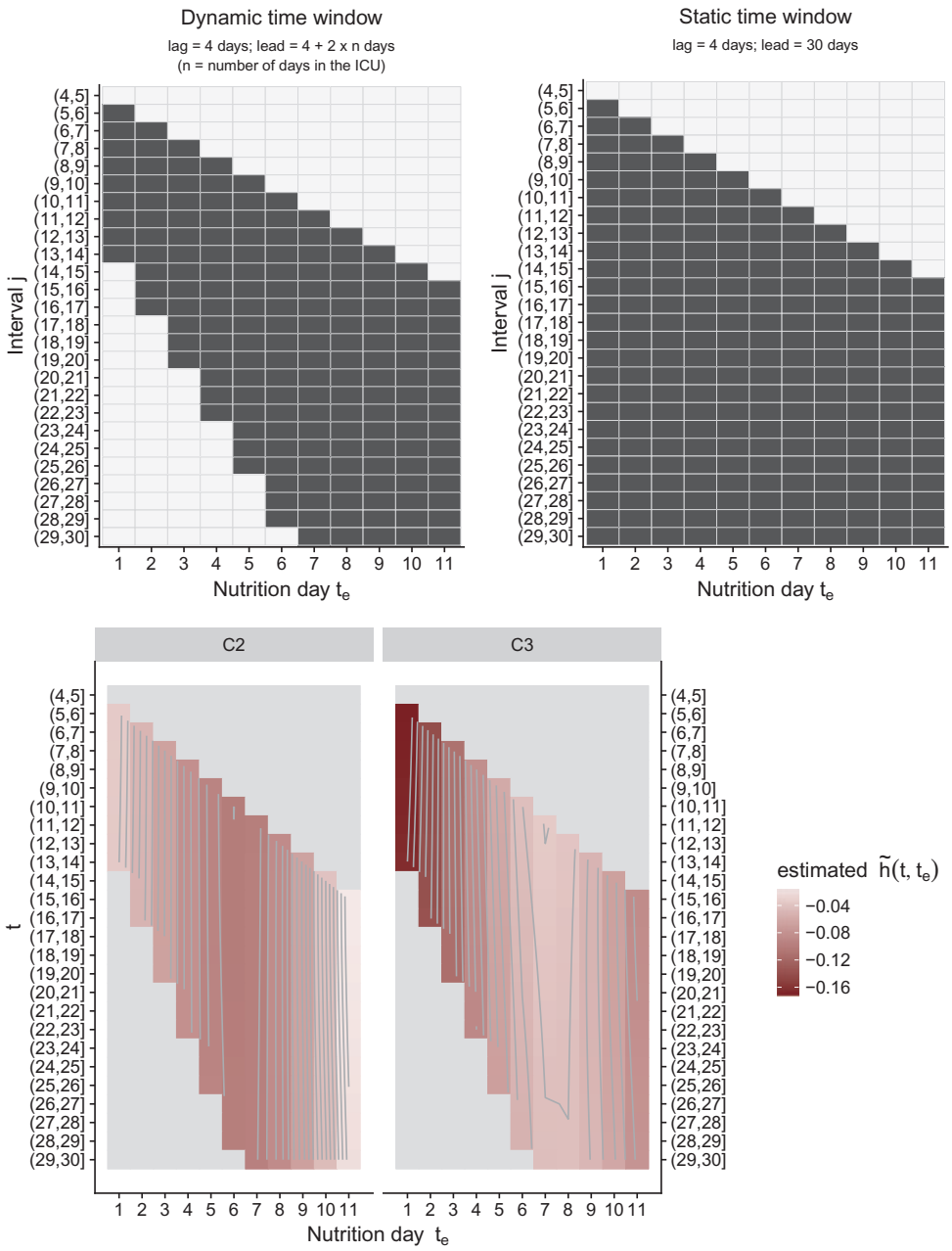


Fig. 1. *Top*: Two possible specifications of the lag-lead window $\mathcal{T}_e(j), j = 5, \ldots, 30$. Left panel shows the *dynamic* $\mathcal{T}_e(j)$ ($t_{\text{lag}} = 4$, $t_{\text{lead}} = t_{\text{lag}} + 2 * t_e$), right panel depicts the *static* $\mathcal{T}_e(j)$ ($t_{\text{lag}} = 4$, $t_{\text{lead}} = 30$) with a longer and constant lead period. This implies that once in effect, partial effects contribute to the cumulative effect until the end of the follow up. *Bottom*: Estimated partial effects surface $\tilde{h}(t, t_e)$ (cf. Equation (3.1)) for nutrition categories $C2$ (left panel) and $C3$ (right panel), respectively. Both estimations are based on the dynamic time window definition (Top, left panel). No estimates are given for values (t, t_e) outside of the window of effectiveness $\mathcal{T}_e(j)$. Estimates can be interpreted as "additive change of the log-hazard at time t if patient is in category $C2$ [$C3$] instead of category $C1$ at exposure time t_e".

while hazards are modeled beginning 96 h after ICU admission for up to 30 days. More generally, prior exposure could be measured over the course of years or decades and survival after hospitalization observed over months or weeks. Non-linear ELRAs based on smooth functions $\tilde{h}(t, t_e, z(t_e))$ are constructed analogously to bivariate smooth functions in 2.2.3 via straightforward extension to three-dimensional tensor products (Wood, 2006, sec. 4.1.8). The complete specification of the model is then given by:

$$\log\left(\lambda_i(t|\mathbf{x}_i, \mathbf{z}_i, \ell_i)\right) = \log(\lambda_0(t)) + \sum_{p=1}^{P} f_p(x_{i,p}, t) + g(\mathbf{z}_i, t) + b_{\ell_i} \tag{2.5}$$

2.3. *Estimation and inference*

Reliable likelihood-based methods for the parameter estimation of the proposed model have been developed in Wood (2011) in the context of penalized models of the form $D(\boldsymbol{\gamma}) + \sum_s \nu_s \boldsymbol{\gamma}' \boldsymbol{K}_s \boldsymbol{\gamma}$, where $D(\boldsymbol{\gamma})$ is the model deviance, $\boldsymbol{\gamma}$ contains all spline basis coefficients and random effects representing model (2.5), and ν_s and $K_s, s = 1, \ldots, S$ are the smoothing parameters and penalty matrices for the individual smooth terms, respectively. In our case, $D(\boldsymbol{\gamma})$ is the deviance of the Poisson GAMM implied by (2.5) in analogy to (2.2). Given $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_S)$, parameter estimates can be obtained by penalized iteratively re-weighted least squares (P-IRLS). To guarantee convergence, Wood (2011) employs P-IRLS based on nested iterations, i.e., after each P-IRLS step, estimation of $\boldsymbol{\nu}$ is updated given the current $\boldsymbol{\gamma}$ estimates. Estimation of $\boldsymbol{\nu}$ proceeds by numerical optimization of the (restricted) maximum likelihood. Subsequent articles (Marra and Wood, 2011, 2012; Wood, 2012) develop shrinkage based procedures for simultaneous smoothness and variable selection and methods for confidence intervals (CIs) and significance tests for penalized model terms. In the following sections, we extend these methods to the context of PAMMs and, particularly, ELRAs. Note that these methods provide valid inference only for estimates derived from a single, pre-specified model. As always, any subsequent model and variable selection based on fits from multiple model specifications would require appropriate adjustments by methods of post selection inference. Without such adjustments, *P*-values and CIs will no longer be valid as estimation uncertainty is likely to be underestimated.

2.3.1. *Confidence intervals.* CIs with good coverage properties for smooth terms are developed in Marra and Wood (2012) and are directly applicable to ELRAs in (2.5). Let $\boldsymbol{\gamma}_z = (\gamma_{1,1}, \ldots, \gamma_{M,K})$ be the vector of estimated basis coefficients in (2.4), and $\mathbf{V}_{\hat{\boldsymbol{\gamma}}_z}$ the empirical Bayesian covariance matrix (Wood, 2006, Ch. 4.8) of their estimates $\hat{\boldsymbol{\gamma}}_z$. Let further $\mathbf{Z}$ be the $J \times MK$ design matrix for a specific exposure history $\mathbf{z}$, where J is the number of intervals into which the follow-up period has been partitioned, and MK is the number of columns associated with the tensor product basis of the ELRA term. The interval-wise CIs of the cumulative effect are given by

$$\mathbf{Z}\hat{\boldsymbol{\gamma}}_z \pm \zeta_{1-\alpha/2}\sqrt{\text{diag}(\mathbf{Z}\mathbf{V}_{\hat{\boldsymbol{\gamma}}_z}\mathbf{Z}^T)} = \hat{\mathbf{g}}_z \pm \zeta_{1-\alpha/2}\widehat{\mathbf{SE}}_z, \tag{2.6}$$

where ζ_q is the q-quantile of the standard normal distribution. In (2.6), $\hat{\mathbf{g}}_z$ is the length J vector estimate of $\mathbf{g}_\mathbf{z} = (g(\mathbf{z}, \tilde{t}_1), \ldots, g(\mathbf{z}, \tilde{t}_J))$, representing the cumulative effect of exposure history $\mathbf{z}$ on the log hazard and $\widehat{\mathbf{SE}}_z$ its standard errors in intervals $j = 1, \ldots, J$. In order to quantify hazard rate differences over time given different exposure histories $\mathbf{z}_{(1)}$ and $\mathbf{z}_{(2)}$, we define $\mathbf{Z} := \mathbf{Z}_{(1)} - \mathbf{Z}_{(2)}$ in (2.6). We can now obtain estimated *differences* $\hat{g}(\mathbf{z}_{(1)}, \tilde{t}_j) - \hat{g}(\mathbf{z}_{(2)}, \tilde{t}_j)$ in cumulative effects for each interval and respective pointwise CIs for these differences given different exposure histories. We demonstrate this approach in Section 3.3 (cf. Figure 2) and investigate properties of such CIs by means of a simulation study in B.1 of supplementary material available at *Biostatistics* online.

2.3.2. *Hypothesis testing.* The method introduced in Section 2.3.1 provides a way to quantify the uncertainty associated with a cumulative effect's estimate at individual time-points of the follow-up and thereby, to asses if the effect differs from zero at certain times t. In some applications, however, it is also of interest to assess the overall effect of an ELRA term. To perform such a global test, we can use significance tests for individual smooth terms of the form $H_0 : \mathbf{f}_q = \mathbf{0}$ (Wood, 2012), where $\mathbf{f}_q$ could be any of the model terms in (2.5) and particularly the ELRA $g(\mathbf{z}, t)$ in 2.4. Using the notation from Section 2.3.1 we define the overall test by

$$H_0 : \mathbf{g_z} = \mathbf{0}. \tag{2.7}$$

The general idea of the test is straightforward and uses the representation of the smooth term as a linear transformation of basis coefficients $\boldsymbol{\gamma}_z$ such that $\mathbf{Z}\boldsymbol{\gamma}_z = \mathbf{g_z}$. An appropriate test-statistic has the familiar quadratic Wald-type form $T_r = \hat{\mathbf{g}}_\mathbf{z}^T \mathbf{V}_{\mathbf{g_z}}^{r^-} \hat{\mathbf{g}}_\mathbf{z}$. Here $\mathbf{V}_{\mathbf{g_z}}^{r^-}$ is the rank r pseudo inverse of $\mathbf{V}_{\mathbf{g_z}} = \mathbf{Z}\mathbf{V}_{\boldsymbol{\gamma}_\mathbf{z}}\mathbf{Z}^\mathbf{T}$. The difficult part then becomes choosing the appropriate r in the context of penalized estimation, as naive choices (e.g., rank of $\mathbf{V}_{\mathbf{g_z}}$) lead to reduced power (see Wood, 2012 for details). Given r, which in this context can be a non-integer number, T_r follows a mixture of χ^2 distributions, from which P-values can be obtained routinely (Wood, 2012, p. 4). In B.1 of supplementary material available at *Biostatistics* online, we show that this *Overall Test* works well when testing individual ELRA terms.

3. Association between caloric intake and mortality in ICU patients

3.1. *Data and objective*

We apply our method in a retrospective analysis of a large international multicenter study with $n = 9661$ critically ill patients (after pre-processing and application of exclusion criteria) with a maximal follow up of 60 days or until release from hospital (cf. Heyland *and others*, 2011). On the day of admission (Day 0), goal calories for each patient were determined by a nutritionist or physician and the actual caloric intake provided by the hospital staff was recorded for a maximum of 11 *calendar* days after the date of ICU admission, which we denote by $t_e \in \{1, \ldots, 11\}$. The follow-up time t was partitioned into 24 h periods starting with ICU admission. We are interested in the association between caloric adequacy and acute mortality, that is, mortality within $t_{\max} = 30$ days (720 h) after ICU admission. In total, 1974 (20.4%) patients died within this period. For our main analysis, we assume that patients released from the hospital survived at least until $t = 30$ (see Discussion in Appendix A.8 of supplementary material available at *Biostatistics* online). We only included patients that survived at least 96 h (4 days), consequently, we began evaluation in interval $(4, 5]$. For the purpose of the analysis presented here, all patients still alive after $t = 30$ were censored.

3.2. *Modeling approach*

We adjusted for the most relevant potential confounders, including subject specific covariates age, body mass index (BMI), sex, diagnosis at admission and admission category, and the Apache II Score (an overall measure of the patients' health status at admission) as well as patient unrelated covariates like year of admission and a random effect (Gaussian frailty) for the ICUs. Since we model the mortality risk beginning in interval $(4, 5]$ (due to application of exclusion criteria), we also included variables that describe the patients' ICU stay up to that point, namely number of days under MV and number of days with additional OI, number of days with parenteral nutrition (PN), and number of days receiving Propofol (PF) on the first three full calendar days of the ICU stay, respectively. To be able to compare different caloric intakes independently of a patient's weight and caloric requirements, we defined a patient's caloric adequacy as $CA = \frac{\text{actual daily caloric intake (in kcal)}}{\text{goal calories (in kcal)}}$. In order to account for unmeasured additional OI, we used a discretized

version of CA, with three categories $C1 : CA < 30\%$, $C2 : 30\% \leq CA < 70\%$, and $C3 : CA \geq 70\%$. If patients received additional OI, they were moved up one category (more details in Appendix A.1 and Table S3 of supplementary material available at *Biostatistics* online).

The effect of nutrition is represented in the model by two ELRAs $g_{C2}(\mathbf{z}_i^{C2}, t)$ and $g_{C3}(\mathbf{z}_i^{C3}, t)$ for dummy variables $z_i^{C2}(t_e), z_i^{C3}(t_e)$ indicating whether nutrition at time t_e was in category $C2$ or $C3$, respectively. Both terms have the structure defined in (2.4). Category $C1$ is the "reference" category, thus direct interpretation of g_{C2} and g_{C3} is only possible with respect to a (hypothetical) patient that received $C1$ on all 11 protocol days. Equation (3.1) shows the model specification:

$$\log(\lambda_i(t|\mathbf{x}_i, \mathbf{z}_i, \ell_i)) = \log(\lambda_0(t)) + \sum_{p=1}^{P} f_p(x_{i,p}, t) + g_{C2}(\mathbf{z}_i^{C2}, t) + g_{C3}(\mathbf{z}_i^{C3}, t) + b_{\ell_i}, \tag{3.1}$$

where $\lambda_0(t)$ represents the baseline hazard rate, $\sum_{p=1}^{P} f_p(x_{i,p}, t)$ incorporates all linear and non-linear effects of time-constant covariates, g_{C2} and g_{C3} are non-linear time-varying cumulative effects of the nutritional intake (see Section 2.2.4 for details), and b_{ℓ_i} is an independent identically distributed Gaussian random intercept term attributed to different ICUs in the data set. Details on model specification and on the estimated confounder effects are in A.3 and A.5 of supplementary material available at *Biostatistics* online. All non-linear functions of time-constant covariates $f_p(\mathbf{x}_{\cdot,p}, t)$ were estimated using P-Splines (Eilers and Marx, 1996) with second order difference penalties and $M = 10$ basis functions on equidistant knots. For the $\tilde{h}(t_e, t)$ terms associated with the ELRA (cf. Equation (2.4)), $M = K = 5$ basis functions were used for each dimension and first order difference penalties were used for exposure time t_e.

The lag-lead window $\mathcal{T}_e(j)$ (see Equation (3.1)) was defined based on substantive considerations with $t_{\text{lag}}(t_e) \equiv 4$ and $t_{\text{lead}}(t_e) = t_{\text{lag}} + 2 \cdot t_e$ and is depicted in Figure 1 (top, left). Viewed column-wise, each panel in Figure 1 shows the intervals at which a specific protocol day ($t_e \in \{1, ..., 11\}$) can *potentially* affect the estimated hazard. Viewed row-wise, the figures show protocol days t_e, which contribute to the cumulative nutrition effect in interval j. We will refer to this specification as *dynamic* lag-lead $\mathcal{T}^{dynamic}$. The clinical considerations underlying this definition are summarized in A.4 of supplementary material available at *Biostatistics* online. To investigate the impact of a possible misspecification of $\mathcal{T}$, we conducted a simulation study (cf. B.1 of supplementary material available at *Biostatistics* online), which closely resembles the data structure and effects of this application example. Section 5 provides discussion on $\mathcal{T}$ and some justification for the choice of t_{lag}.

3.3. *Results*

We are mostly interested in the relationship between caloric intake and survival. Therefore, in the following, we will only present the results of this association. Estimated confounder effects are presented in A.5 of supplementary material available at *Biostatistics* online. The estimated partial effect surfaces $\tilde{h}_{C2}(t_e, t)$ and $\tilde{h}_{C3}(t_e, t)$ for cumulative effects g_{C2} and g_{C3}, respectively, are presented at the bottom of Figure 1 (see also Figure S5 of supplementary material available at *Biostatistics* online for CIs). Both effects are estimated to be almost constant over t (vertical axis) and vary meaningfully only over t_e (horizontal axis). Estimates can be interpreted as the estimated difference in log-hazard rates at time t between patients who were in category $C2$ ($C3$) at time t_e compared to a patient who was in category $C1$ at time t_e if the two had identical exposures at all other exposure time-points $t_e' \neq t_e$ and identical values for random effects and all other covariates $x_p, p = 1, \ldots, P$. For example, a patient with $C2$ instead of $C1$ at $t_e = 7$ is estimated to have a decreased hazard by a factor of about $e^{-0.1} \approx 0.90$ over $t \in (11, 30]$, compared with a patient with identical covariate values and otherwise identical exposure history. Direct interpretation of these partial effects, however, is hindered by the fact that hazard at time t is affected by different numbers of exposures depending on the structure of $\mathcal{T}_e(j)$, since the total effect of exposure history $\mathbf{z}$ on the hazard is given by

Table 2. *Overview of evaluated comparisons with nutrition categories C1 (lower), C2 (mid) and C3 (upper) as defined in Table S3 supplementary material available at Biostatistics online*

Comparison	$\mathbf{z}_{(1)}$	$\mathbf{z}_{(2)}$
Comparison A	Days 1–11: $C1$	Days 1–4: $C1$, Days 5–11: $C2$
Comparison B	Days 1–11: $C1$	Days 1–11: $C2$
Comparison C	Days 1–4: $C1$, Days 5–11: $C2$	Days 1–11: $C2$
Comparison D	Days 1–11: $C1$	Days 1–11: $C3$
Comparison E	Days 1–11: $C2$	Days 1–4: $C2$, Days 5–11 $C3$
Comparison F	Days 1–11: $C2$	Days 1–11: $C3$

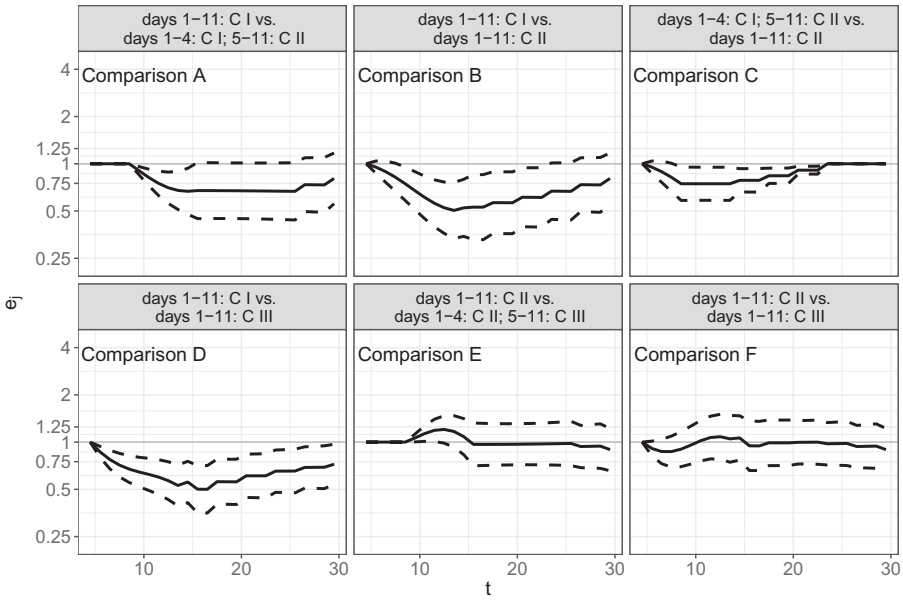


Fig. 2. Estimated hazard ratios $\hat{e}_j$ ($\pm 2\widehat{SE}_{\hat{e}_j}$) for the comparison of different nutrition protocols (all other covariates being equal), as defined in equation (3.2). Each facet depicts one comparison. The comparisons are indicated in the facet headers and summarized in Table 2. $\hat{e}_j < 1$ indicate an increase in the hazard rate for the first protocol compared with the second. Dashed lines depict approximate 95% CI as described in Section 2.3.1.

$g_{C2}(\mathbf{z}^{C2}, t) + g_{C3}(\mathbf{z}^{C3}, t)$, and since this effect represents the difference in the log-hazard compared with a patient with constant category $C1$ nutrition. For these reasons, we analyze and interpret estimated hazard ratios between hypothetical patients with different clinically relevant exposure histories that can easily be computed from the partial effects. Therefore, we show the cumulative effects of nutrition, evaluated at interval midpoints $\tilde{t} \in \{4.5, 5.5, \ldots, 29.5\}$, as hazard ratios

$$e_j = \frac{\lambda(\tilde{t}_j|\mathbf{z}_2)}{\lambda(\tilde{t}_j|\mathbf{z}_1)} = \exp(\mathbf{g}_{\mathbf{z}_2} - \mathbf{g}_{\mathbf{z}_1}) \tag{3.2}$$

for patients with different nutrition protocols $\mathbf{z}_1$ and $\mathbf{z}_2$ and identical values for all other covariates. Six clinically relevant comparisons are summarized in Table 2. The results are depicted in Figure 2 and suggest

that (i) hypocaloric (category $C1$) nutrition is associated with increased hazard rates throughout the follow-up period (Comparisons B, D and to a lesser extent Comparison A); (ii) based on this model, moving from constantly medium ($C2$) to constantly full ($C3$) nutrition is not associated with a decrease of the hazard rate (Comparisons E, F); (iii) the (small) hazard rate increases associated with hypocaloric nutrition in the first few days of the ICU stay may persist for up to 25 days after ICU admission (Comparison C).

In A.7 supplementary material available at *Biostatistics* online, we compare our main model with two alternative models, which only use time-aggregated cumulative nutrition information as TDCs instead of fitting cumulative effects of the nutrition exposure histories. The findings suggest that, for these data, results qualitatively similar to the ones presented here could have been obtained using simpler approaches. We argue, however, that the suggested approach is preferable: Since the true association structure is typically not known in advance and since the penalized estimation of our very flexible proposal seems to be able to avoid overfitting in settings where the association has a simple linear or time-constant structure. Allowing for complex associations entails little cost and helps to avoid both possible model misspecifiation and issues of post-selection inference that come with *post hoc* comparisons of simpler models using different pre-specified variations of cumulating the (effects of) TDCs (cf. A.7 of supplementary material available at *Biostatistics* online for details).

4. SIMULATION STUDY

We performed extensive simulation studies to investigate the performance of the proposed modeling approach. As general performance of this model class has already been evaluated in other publications (cf. Wood, 2011 for GAMs and Argyropoulos and Unruh, 2015 for their application to survival analysis), our simulation focuses on the evaluation of ELRAs and is divided in two parts. In *Part A* (Section 4.2), we demonstrate the ability of our approach to estimate the effects of a complex nonlinear DLNM as well as a time-varying extension thereof, i.e., a three-dimensional function of the form $g(\mathbf{z}, t) = \int h(t, t - t_e, z(t_e))\mathrm{d}t_e$.

Simulation *Part B* is modeled after the application example, especially with respect to data structure and the structure, magnitude, and shape of simulated effects and focuses on estimation of cumulative effects $g(\mathbf{z}, t)$ instead of partial effects $h(t, t_e, z(t_e))$. Details on data generation and results of Simulation Part B are provided in B.1 of supplementary material available at *Biostatistics* online. They show that simulated effects are generally estimated well and CIs as well as hypothesis tests discussed in Sections 2.3.1 and 2.3.2 have good properties, often even when the model is misspecified (cf. B.1.2 and Figures S19, S20 of supplementary material available at *Biostatistics* online).

4.1. *Data generation*

In both simulation parts, random survival times were drawn from the piece-wise exponential distribution. To do so, one needs to specify a vector of piece-wise constant hazards $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_J)$ in intervals defined by time-points $\boldsymbol{\kappa} = (\kappa_0, \ldots, \kappa_J)$. In general, $\boldsymbol{\lambda}$ can be defined through a function of time t, covariates $\mathbf{x}$ and exposure histories $\mathbf{z}$ such that $\lambda_j, j = 1, \ldots, J$ is given by $\lambda(t|\mathbf{x}, \mathbf{z}) = f(t, \mathbf{x}, \mathbf{z})$, evaluated at times $t = \kappa_j, j = 1, \ldots, J$. The algorithm by which survival times are drawn from the piece-wise exponential distribution ($t \sim \mathrm{PEXP}(\boldsymbol{\lambda}, \boldsymbol{\kappa})$) is described in detail in B of supplementary material available at *Biostatistics* online, specifically Table S7. Note that drawing from PEXP generates continuous event times, thus for each subject we obtain tuples $(t_i, x_i, \mathbf{z}_i)$.

4.2. *Simulation Part A*

We adapt a simulation study performed in Gasparrini *and others* (2017) for Poisson responses to the context of time-to-event data. Our objective is to demonstrate the ability of our method to estimate DLNMs and

time-varying DLNMs. We distinguish two scenarios, both of which simulate data from a model with a single cumulative effect of a TDC.

Scenario (1): Simulate a "classical" DLNM for survival data via

$$\log(\lambda(t, \mathbf{z})) = \log(\lambda_0(t)) + \int h(t - t_e, z(t_e))\mathrm{d}t_e \tag{4.1}$$

Scenario (2): Extend scenario (1) such that the ELRA additionally varies over time

$$\begin{aligned}\log(\lambda(t, \mathbf{z})) &= \log(\lambda_0(t)) + g(\mathbf{z}, t) = \log(\lambda_0(t)) + \int \tilde{h}(t, t - t_e, z(t_e))\mathrm{d}t_e \\ &= \log(\lambda_0(t)) + \int f(t) \cdot h(t - t_e, z(t_e))\mathrm{d}t_e \end{aligned} \tag{4.2}$$

For the functional shape of $h(t - t_e, z(t_e))$ in (4.1) and (4.2) we used the "complex" ELRA, which is depicted at the right panel ("TRUTH") in Scenario (1) in Figure 3 (see also Figure 1 in Gasparrini *and others*, 2017). With Φ_{μ,σ^2} the density function of the normal distribution with mean μ and variance σ^2, the mathematical representation of h is given by the functions $f_1(z(t_e)) = \Phi_{1.5,2}(z(t_e)) + 1.5 \cdot \Phi_{7.5,1}(z(t_e))$, $f_2(t - t_e) = 15 \cdot \Phi_{8,10}(t - t_e)$, $f_3(t - t_e) = 5 * (\Phi_{4.6}(t - t_e) + \Phi_{25,4}(t - t_e))$ and finally $h(t - t_e, z(t_e)) = \begin{cases} f_1(z(t_e)) \cdot f_3(t - t_e) & \text{if } z(t_e) \geq 5 \\ f_1(z(t_e)) \cdot f_2(t - t_e) & \text{if } z(t_e) \geq 5 \end{cases}$ (cf. Appendix in Gasparrini *and others*, 2017).

In (4.2), we set $f(t) = -\cos(\pi t / t_{\max})$, such that the partial effects are negative at the beginning and then become positive after the first half of the follow up (cf. right panel ("TRUTH") of Scenario (2) in Figure 3). In both Scenarios, in addition to fitting a correctly specified model to the simulated data, we also fit a underspecified model for comparison. In Scenario (1), we fit a WCE with partial effects $h(t - t_e)z(t_e)$, thus neglecting the non-linearity in $z(t_e)$. In Scenario (2), we fit a standard (non time-varying) DLNM with partial effects $h(t - t_e, z(t_e))$, thus neglecting the additional time-variation. All models were estimated by PAMMs.

Evaluation was performed by graphical comparison of the estimated $\hat{h}_{t,t_e,z(t_e)} \equiv \hat{h}(t - t_e, z(t_e))$ [Scenario 2: $\hat{h}(t, t - t_e, z(t_e))$] to the respective true (simulated) surface $h_{t,t_e,z(t_e)}$ and with summary statistics $RMSE = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{400}\sum_{t-t_e=0}^{40}\sum_{z(t_e)=0}^{10}(h_{t,t_e,z(t_e)} - \hat{h}_{t,t_e,z(t_e)})^2}$ and $\text{coverage}_\alpha = \frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{400}\sum_{t-t_e=0}^{40}\sum_{z(t_e)=0}^{10} I\left(h_{t,t_e,z(t_e)} \in [\hat{h}_{t,t_e,z(t_e)} \pm \zeta_{1-\alpha/2}\hat{\sigma}_{\hat{h}_{t,t_e,z(t_e)}}]\right)\right]$, where ζ_q is the q-quantile of the standard normal distribution and $\hat{\sigma}_{\hat{h}_{t,t_e,z(t_e)}}$ is the estimated standard deviation of the partial effect estimate. In Scenario (2) RMSE and coverage are additionally averaged over each time-point $t = 1, \ldots, 40$.

The results for Scenario (1) are shown at the top of Figure 3. The simulated effects are estimated very well when the model is specified correctly—"PAM (DLNM)" has a mean RMSE of ≈ 0.037 and close to nominal coverage of $\approx 97\%$. For comparison, the misspecified model "PAM (WCE)", which wrongly assumes linearity in $z(t_e)$, yields an 1.6 times increased RMSE (≈ 0.06) and a coverage of $\approx 36\%$. The results for Scenario (2) are presented at the bottom of Figure 3. The tri-variate function $h(t, t - t_e, z(t_e))$ was not estimated quite as precisely by the correctly specified model ("PAM (TV DLNM)") as in the time-constant Scenario (1). In general, however, the ELRA, as well as its variation over time were fit very well, which illustrates the ability of our approach to estimate truly three-dimensional cumulative effects (4.2) (mean RMSE of about 0.024 and mean coverage of about 0.96). The simpler model ("PAM (DLNM)"), that ignores possible time-variation, consequently has the same value at all time-points and is close to zero, i.e., the estimate averages out the risk reducing effects at the beginning and the risk

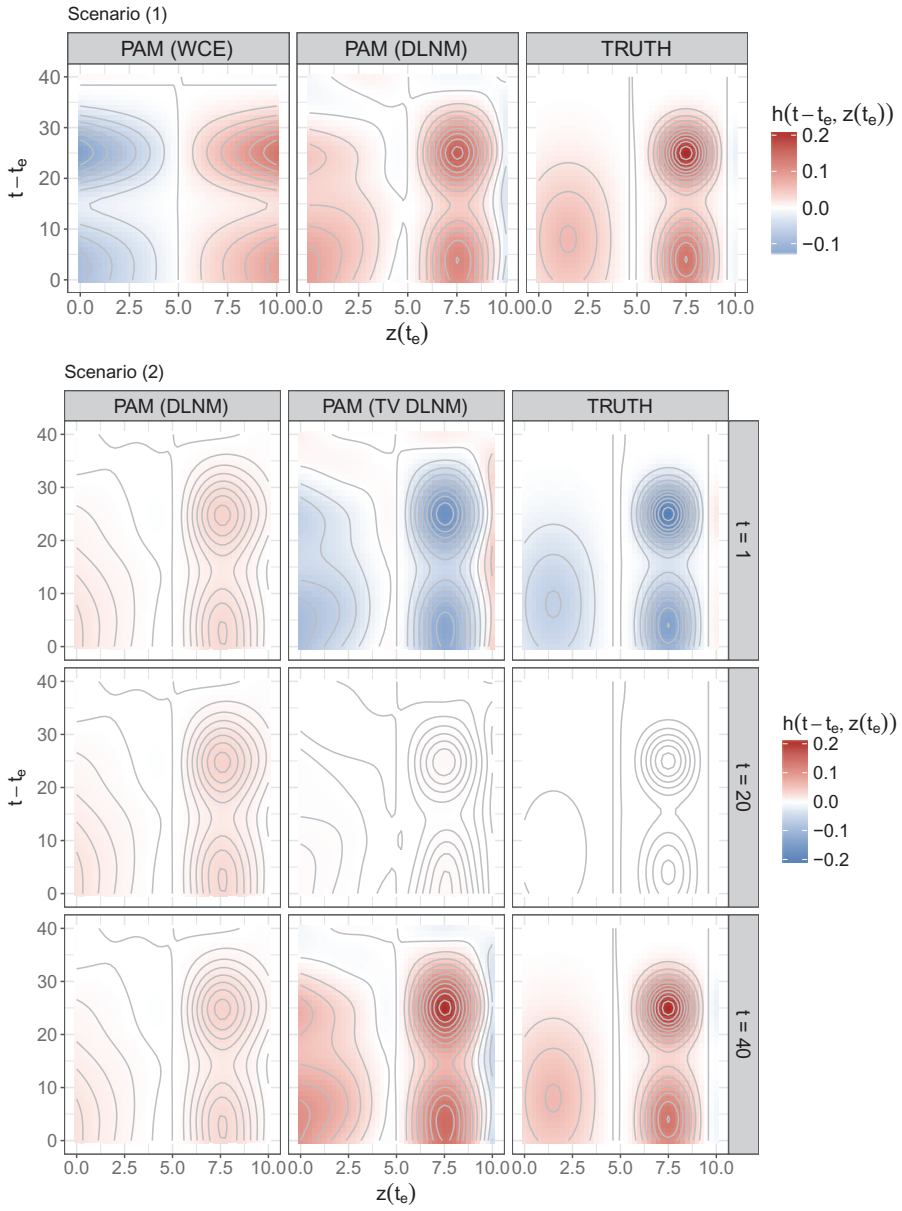


Fig. 3. Simulation Part A *Scenario (1)*: estimated WCE ($\hat{h}(t-t_e)z(t_e)$; left), estimated DLNM ($\hat{h}(t-t_e, z(t_e))$; middle), and true simulated bi-variate effect surfaces ($h(t-t_e, z(t_e))$; right) for different values of latency $t-t_e$ and exposure $z(t_e)$. *Scenario (2)*: estimated (left and mid panel) and true simulated (right panel) bi-variate effect surfaces $\hat{h}(t-t_e, z(t_e))$, and $h(t-t_e, z(t_e))$, respectively, for different values of latency $t-t_e$, exposures $z(t_e)$ and for follow-up times $t \in \{1, 20, 40\}$. In both Scenarios, displayed estimates are obtained by averaging over all estimates from $R = 500$ simulation runs.

increasing effects at the end of the follow-up, respectively (RMSE ≈ 0.038, coverage $\approx 71\%$). Note that RMSEs for Scenario (2) are generally smaller, due to small effects and estimates toward the middle of the follow-up.

5. Discussion

By embedding the concept of PEMs into the framework of additive models, we define a very versatile model class for life-time data analysis that inherits the robust and flexible tools for modeling, estimation and validation available for GAMMs. In contrast to traditional PEMs, the baseline and time-varying effects are represented as flexible, potentially non-linear penalized splines. Our general approach to flexibly modeling cumulative effects of TDCs (ELRAs: exposure-lag-response associations) takes into account the entire exposure history, i.e., both the timing and amount of exposure. The proposed presentation of ELRAs in form of hazard ratio trajectories for pairs of realistic exposure histories provides an accessible alternative to established visualization techniques like contour or wire-frame plots.

The general formulation, we propose defines a broad model class that includes previous approaches for cumulative effects like the WCE model (Sylvestre and Abrahamowicz, 2009) and the DLNM (Gasparrini *and others*, 2017) as special cases. The application example (Section 3) is a direct demonstration of this generalization, and—although the improvement in predictive accuracy was small since the estimated partial effects did not vary much over t in this case — also provides an example for the estimation of a more flexible WCE model, as the weight function $h(t, t_e)$ was allowed to depend on the specific combination of t and t_e instead of restricting the dependency to the latency $t - t_e$ a priori. Simulation study Part B (cf. Section 4.2) additionally illustrates the practical feasibility of a "time-varying" extension of the DLNM.

In general, our simulation studies (Section 4) confirm that our method is able to estimate complex ELRAs and is relatively robust to misspecification. When no true exposure effect is present, both the coverage of the proposed CIs for all comparisons and the Type I error rate of the hypothesis tests maintained near nominal levels, regardless of the specification of the penalty and the correctness of the specified lag and lead times (cf. supplement Figures S19 (upper panel, Setting IV) and S20), but CIs can sometimes have sub-nominal coverage for misspecified non-null models, especially in the case of P_1 penalties. It is also apparent that a misspecification of lag and lead times can induce bias, potentially causing an underestimation of effects at the end of the follow up (if the lead time was specified as too short), or an overestimation (if the lead time was specified as too long). Going forward, data driven selection of the relevant time window is our most important research goal. One approach for such a procedure could include additional penalties similar to the double penalty approach by Obermeier *and others* (2015) for distributed lag models.

From a practical perspective, the interpretation of (cumulative) effects of TDCs is problematic whenever the exogeneity of such variables is unclear. For example, although nutrition is administered by the hospital staff, the amount of nutrition provided is likely to depend on patients' health status, at least to some degree: patients undergoing procedures due to life-threatening complications presumably receive less calories or feeding could be stopped due to the decision to withdraw life support. Handling such variables always requires a trade-off with respect to the recency of the covariate (Crowder, 2012, Ch. 3.6.) that may result in better adjustment for confounding for more recent values of the covariate, but may also be fully indicative of the outcome and thus induce indication bias (Signorello *and others*, 2002). In our application, we address this issue by including a minimum lag time of 4 days for the nutritional effects. In future research, our method could be combined with frameworks that take into account that TDCs can simultaneously be confounders and mediators of past exposures. Xiao *and others* (2014) recently presented one such extension of the WCE to marginal structural models. Additionally, as discussed in Supplement A.8, extending the proposed estimation of cumulative effects to the competing risks setting, similar to Danieli and Abrahamowicz (2017), would also be of great value for practical applicability.

Finally, the flexibility of this model class implies challenges with respect to variable and model selection. Wynant and Abrahamowicz (2014) address the issue of model selection for non-linear and/or time-varying effects in the context of partial likelihood models, and they demonstrate the usefulness of backward elimination for this purpose. Boosting based approaches, which have been shown to perform well for similarly complex GAMs in the context of functional data analysis and for complex time-to-event models might also be of interest here.

6. SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org. Data and code to reproduce the analyses are available at https://github.com/adibender/elra-biostats.

ACKNOWLEDGMENTS

The authors would like to thank Daren Heyland for providing the data set and useful discussion. We also want to thank the referees for their detailed and helpful comments, which greatly improved the manuscript. *Conflict of Interest*: None declared.

FUNDING

Fabian Scheipl was supported by the German Research Foundation through the Emmy Noether Programme (GR 3793/1-1 to Sonja Greven).

REFERENCES

ARGYROPOULOS, C. AND UNRUH, M. L. (2015). Analysis of time to event outcomes in randomized controlled trials by generalized additive models. *PLoS One* **10**, e0123784.

BERHANE, K., HAUPTMANN, M. AND LANGHOLZ, B. (2008). Using tensor product splines in modeling exposure-time-response relationships: application to the colorado plateau uranium miners cohort. *Statistics in Medicine* **27**, 5484–5496.

CROWDER, M. J. (2012). *Multivariate Survival Analysis and Competing Risks*. Chapman & Hall/CRC Texts in Statistical Science. Hoboken: CRC Press.

DANIELI, C. AND ABRAHAMOWICZ, M. (2017). Competing risks modeling of cumulative effects of time-varying drug exposures. *Statistical Methods in Medical Research*, doi:10.1177/0962280217720947.

DEMARQUI, F. N., LOSCHI, R. H. AND COLOSIMO, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis* **14**, 333–356.

EILERS, P. H. C. AND MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121.

FRIEDMAN, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics* **10**, 101–113.

GASPARRINI, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine* **33**, 881–899.

GASPARRINI, A., SCHEIPL, F., ARMSTRONG, B. AND KENWARD, M. G. (2017). A penalized framework for distributed lag non-linear models. *Biometrics* **73**, 938–948.

GERDS, T. A., KATTAN, M. W., SCHUMACHER, M. AND YU, C. (2013). Estimating a time-dependentâĂL'concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine* **32**, 2173–2184.

HASTIE, T. AND TIBSHIRANI, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)* **55**, 757–796.

HEYLAND, D. K., CAHILL, N. AND DAY, A. G. (2011). Optimal amount of calories for critically ill patients: depends on how you slice the cake! *Critical Care Medicine* **39**, 2619–2626.

MARRA, G. AND WOOD, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis* **55**, 2372–2387.

MARRA, G. AND WOOD, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics* **39**, 53–74.

OBERMEIER, V., SCHEIPL, F., HEUMANN, C., WASSERMANN, J. AND KÜCHENHOFF, H. (2015). Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**, 395–412.

SIGNORELLO, L. B., MCLAUGHLIN, J. K., LIPWORTH, L., FRIIS, S., SØRENSEN, H. T. AND BLOT, W. J. (2002). Confounding by indication in epidemiologic studies of commonly used analgesics. *American Journal of Therapeutics* **9**, 199–205.

SYLVESTRE, M.-P. AND ABRAHAMOWICZ, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine* **28**, 3437–3453.

WHITEHEAD, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **29**, 268–275.

WOOD, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: CRC Press.

WOOD, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 3–36.

WOOD, S. N. (2012). On p-values for smooth components of an extended generalized additive model. *Biometrika* **100**, 221–228.

WOOD, S. N., SCHEIPL, F. AND FARAWAY, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing* **23**, 341–360.

WYNANT, W. AND ABRAHAMOWICZ, M. (2014). Impact of the model-building strategy on inference about nonlinear and time-dependent covariate effects in survival analysis. *Statistics in Medicine* **33**, 3318–3337.

XIAO, Y., ABRAHAMOWICZ, M., MOODIE, E. E. M., WEBER, R. AND YOUNG, J. (2014). Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: Reassessing the cardiovascular risks of didanosine treatment in the Swiss HIV cohort study. *Journal of the American Statistical Association* **109**, 455–464.

[*Received February 6, 2017; revised January 15, 2018; accepted for publication January 16, 2018*]

Chapter 4

Calorie intake and short-term survival of critically ill patients

Chapter 4 provides a reanalysis of the data in Chapter 3, with more focus on medical aspects of the analysis and its implications for clinical practice. The article also includes sensitivity and subgroup analyses.

Contributing article:

Hartl, W. H., Bender, A., Scheipl, F., Kuppinger, D., Day, A. G., and Küchenhoff, H. (2018). Calorie intake and short-term survival of critically ill patients. *Clinical Nutrition*.

Author contributions:

Wolfgang W. Hartl, Fabian Scheipl, Andreas Bender and Helmut Küchenhoff developed the statistical analysis strategy. Andreas Bender and Fabian Sheipl performed the statistical modeling and inference. Andreas Bender and Andrew G. Day were responsible for data management and created descriptive statistics and visualizations. Wolfgang Hartl and David Kuppinger did the literature search and drafted the manuscript. All authors contributed to interpretation of the data and critical revision of the manuscript.

Supplementary material available at:

https://doi.org/10.1016/j.clnu.2018.04.005

ARTICLE IN PRESS

Clinical Nutrition xxx (2018) 1–8

Contents lists available at ScienceDirect

Clinical Nutrition

journal homepage: http://www.elsevier.com/locate/clnu

Original article

Calorie intake and short-term survival of critically ill patients

Wolfgang H. Hartl [a, *, 1], Andreas Bender [b, 1], Fabian Scheipl [b], David Kuppinger [a], Andrew G. Day [c], Helmut Küchenhoff [b]

[a] Department of General, Visceral, Transplantation, and Vascular Surgery, University School of Medicine, Grosshadern Campus, Ludwig-Maximilians-Universität, Munich, Germany
[b] Statistical Consulting Unit, StaBLab, Department of Statistics, Ludwig-Maximilians-Universität, Munich, Germany
[c] Clinical Evaluation Research Unit, Kingston General Hospital, Kingston, Ontario, Canada

ARTICLE INFO

Article history:
Received 30 October 2017
Accepted 6 April 2018

Keywords:
Critical care
Nutrition
Caloric supply
Survival

SUMMARY

Background & aims: The association between calorie supply and outcome of critically ill patients is unclear. Results from observational studies contradict findings of randomized studies, and have been questioned because of unrecognized confounding by indication. The present study wanted to re-examine the associations between the daily amount of calorie intake and short-term survival of critically ill patients using several novel statistical approaches.

Methods: 9661 critically ill patients from 451 ICUs were extracted from an international database. We examined associations between survival time and three pragmatic nutritional categories (I: <30% of target, II: 30–70%, III: >70%) reflecting different amounts of total daily calorie intake. We compared hazard ratios for the 30-day risk of dying estimated for different hypothetical nutrition support plans (different categories of daily calorie intake during the first 11 days after ICU admission). To minimize indication bias, we used a lag time between nutrition and outcome, we particularly considered daily amounts of calorie intake, and we adjusted results to the route of calorie supply (enteral, parenteral, oral).

Results: 1974 patients (20.4%) died in hospital before day 30. Median of daily artificial calorie intake was 1.0 kcal/kg [IQR 0.0–4.1] in category I, 12.3 kcal/kg [9.4–15.4] in category II, and 23.5 kcal/kg [19.5–27.8] in category III. When compared to a plan providing daily minimal amounts of calories (category I), the adjusted minimal hazard ratios for a delayed (from day 5–11) or an early (from day 1–11) mildly hypocaloric nutrition (category II) were 0.71 (95% confidence interval [CI], 0.54 to 0.94) and 0.56 (95% CI, 0.38 to 0.82), respectively. No substantial hazard change could be detected, when a delayed or an early, near target calorie intake (category III) was compared to an early, mildly hypocaloric nutrition.

Conclusions: Compared to a severely hypocaloric nutrition, a mildly hypocaloric nutrition is associated with a decreased risk of death. In unselected critically ill patients, this risk cannot be reduced further by providing amounts of calories close to the calculated target.

Study registration: ID number ISRCTN17829198, website http://www.isrctn.org.

1. Introduction

Patients requiring acute organ support after ICU admission are candidates for artificial nutrition. Appropriate nutrition delivery in the acute phase of critical illness, however, is currently highly controversial. Results of controlled studies have been conflicting, and two major studies have suggested that even a severe energy deficit (daily intake < 600 kcal) may be tolerated during the first week of critical illness [1,2]. These results have been heavily criticized being possibly valid only for highly selected subgroups of critically ill patients [3]. Observational studies, which may be less

Abbreviations: ICU, Intensive Care Unit; HR, Hazard Ratio; CI, Confidence Interval; BMI, Body Mass Index; kg, Kilogram; OI, Oral Intake; EN, Enteral Nutrition; PN, Parenteral Nutrition; MV, Mechanical Ventilation.

* Corresponding author. Klinik für Allgemeine, Viszeral-, Transplantations-, und Gefäßchirurgie, der Universität, Campus Grosshadern, LMU München, Marchioninistr. 15, D-81377 Munich, Germany. Fax: +49 89 4400 4418.
E-mail address: whartl@med.uni-muenchen.de (W.H. Hartl).

[1] W. Hartl and A. Bender contributed equally to this work.

https://doi.org/10.1016/j.clnu.2018.04.005
0261-5614/

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

selective in terms of patient inclusion criteria have yielded contradictory results, and have also been subject to substantial criticism because of inherent methodological weaknesses (e.g., confounding by indication) [4,5]. Correspondingly, active treatment strategies currently recommended in specific guidelines differ to the extent that, during the acute phase of a disease, a higher energy supply is recommended in Europe and Canada, whereas U.S. guidelines and recent international sepsis guidelines largely favor a hypocaloric nutrition [6–9].

To better understand the importance of different levels of calorie intake for short-term survival we analyzed a large international database. Novel statistical techniques were used to minimize limitations of conventional ways of assessing nutritional effectiveness in observational studies (indication bias).

2. Methods

2.1. Study overview

2.1.1. Database

The present study is an analysis of a subset of a large international point prevalence survey of nutrition practice in ICUs (www.criticalcarenutrition.com/ins) conducted in 2007, 2008, 2009 and 2012. Details of the survey are provided in the Supplementary Appendix and elsewhere [10].

Participating ICUs were recruited by disseminating of study information to membership registries of critical care and clinical nutrition societies, and by e-mailing individual health care providers. Each year participating ICUs were asked to enroll 20 consecutive intubated adult (≥18) patients who were ventilated within the first 48 h in the ICU and remained in the ICU for at least 72 h. In total, 12,565 patients from 451 ICUs had been included into the database. 9661 critically ill patients could be identified as fulfilling the inclusion criteria for the current analysis.

Clinical management of study patients was left to the discretion of the treating physician. Shortly after ICU admission, a daily caloric target was determined. Ways to calculate this target were also left to the judgment of the individual provider. The local institutional review boards approved the retrospective anonymous data analysis.

2.1.2. Data collection

Using a secure web-based data collection tool, the following information was collected: date of ICU admission, admission category (elective surgery, emergency surgery, medical), primary admission diagnosis, sex, age, BMI, duration of mechanical ventilation, and Apache II score on admission day. Treating physicians recorded daily the type (oral, enteral, parenteral) and amount of calories, amino acids or protein received from parenteral nutrition (PN), enteral nutrition (EN) or propofol. Daily nutritional variables were collected from the day of ICU admission (partial day) to a maximum of additional 11 days after admission date. In the current analysis, we ignored nutrition received on the day of ICU admission and refer to the subsequent discrete calendar days as "*nutrition days 1 to 11*", while a patient's continuous survival time was calculated as "*days after ICU admission*", where each day described a 24 h period starting on the exact date and time of ICU admission. Registering of calorie intake was stopped before nutrition day 11, if a patient died or was discharged from the ICU before that day. For the first three nutrition days, we registered the number of days on which a patient had been mechanically ventilated, had received PN or propofol, or had been fed orally. Patients were followed for a maximum of 60 days. In this study, we investigate short term survival of 30 days. Patients alive for more than 30 days were censored.

2.2. Patients

2.2.1. Inclusion and exclusion criteria

Patients extracted from the database were ≥18 years of age. They had been treated in an ICU for at least 96 h and had, therefore, a higher chance of benefiting from specific metabolic interventions [11]. In addition, on at least one day during the first 96 h of their ICU stay, patients had to have received artificial (enteral or parenteral) nutrition.

2.2.2. Quantifying calorie intake

To account for unquantified, additional energy intake through oral feeding, total daily calorie intake was classified by using a pragmatic approach defining three different categories based on the amount of received calories. For categorization, we first calculated daily artificial energy intake by summing up calories from enteral nutrition, parenteral nutrition or pharmacological (propofol) supply. If patients had not been fed orally on a certain nutrition day, their total daily energy intake was then expressed as a fraction of the daily caloric target calculated at ICU admission. Finally, a specific nutritional category was assigned to this fraction (category I: <30% of target, category II: 30–70% of target, category III: >70% of target).

If there had been additional oral energy intake on a certain nutrition day, classification of total daily energy intake considered both oral and artificial nutrition in a qualitative manner: first, artificial energy intake was categorized as described above. If there had been any oral intake on a certain day, and if patients had simultaneously received <30% of calculated caloric target by artificial (enteral and/or parenteral) nutrition on that day, total daily energy intake was then classified as belonging to category II; if patients had received 30–70% of calculated caloric target on that day in combination with oral intake, total daily energy intake was assumed to be in category III. Total calorie intake remained in category III, if there had been oral intake in combination with >70% of caloric target. Furthermore, we assumed that patients, who had been discharged alive from the ICU before nutrition day 11, subsequently (up to nutrition day 11) had a daily calorie intake in the range of category III.

2.3. Statistics

Full details of the statistical methods are provided in the Supplementary Appendix and in a separate publication [12]. We performed a survival analysis based on a novel combination of generalized additive models and piecewise exponential models to estimate hazard rates for death beyond day 4 after ICU admission [13]. Our model included several confounder variables, including the use of oral, enteral or parenteral nutrition. For the primary analysis, we assumed that daily oral intake had contained a relevant amount of calories, and that patients discharged from the hospital before day 30 after ICU admission had survived up to this day (best case scenario). Two sensitivity analyses checked the validity of these assumptions. When analyzing associations with survival we also used a time lag of 4 days (*lag-time*) to minimize the indication bias originating from possible changes of calorie intake just prior to death.

To facilitate the interpretation of the associations between caloric supply and outcome, we constructed five different hypothetical nutrition support plans reflecting different levels of daily energy intake during nutrition days 1–11 (Table 1). We designed six different pairwise comparisons of these plans analyzing the time-varying mortality hazard ratios associated with these nutrition support plans in comparison to each other (Figs. 1 and 2, graphs shown on the left). These nutrition support plans represent

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

hypothetical/possible concepts similar to established nutrition protocols, and they do not reflect different patient populations.

We also performed explorative analyses of several subgroups (Apache II Score > 25; BMI ≥ 25 and <25; medical and surgical (elective + emergency) patients).

The statistical programming environment R [14] was used for visualization and data analysis. The models were estimated using the R package mgcv.

3. Results

3.1. Study participants

Nutritional therapies were assessed in 9661 patients who met our inclusion criteria. Of these patients, 1974 patients (20.4%) died in hospital after 96 h but within 30 days of ICU admission. Demographic and clinical characteristics of the patients are listed in Tables 2 and 3. After 30 days, 2981 (30.9%) of the patients were still hospitalized. Discharged patients spent 10 days (Median) (IQR 6.6–15.0 days) in the ICU, 16.3 days (IQR 11.3–22.3 days) in the hospital, and received mechanical ventilation for 6.8 days (Median).

3.2. Nutritional therapy

On at least one nutrition day, 7015 of 9661 patients (72.6%) received very low amounts of calories (<30% of target and no additional oral intake, category I). 7111 patients (73.6%) could not be fed orally during day 1–11 of their ICU stay. Initially, records for 90,898 days of nutritional therapy had been available for the analysis (on average, 9.4 days per patient). Including assumptions on calorie intake after ICU discharge, 102,686 days of nutritional therapy were included into the analysis. On 18,757 nutrition days (18.3%), patients had received less than 30% of the caloric target calculated at ICU admission, and had no additional oral intake; on 21,634 nutrition days (21,1%) calorie intake had been in the range of category II (<30% of target with additional oral intake, or 30–70% of target with no additional oral intake), and on 62,295 nutrition days (60.1%) calorie intake had been in the range of category III (>70% of target regardless of oral intake, or 30–70% of target with additional oral intake).

On the basis of data from nutrition days on which patients had not been fed orally, it was possible to calculate the achieved percentages of daily caloric targets for categories I, II and III, and the precise amount of enteral and/or parenteral calories administered daily (category I: median 4.3% of target [IQR, 0.0–17.8], 1.0 kcal/kg day [0.0–4.1]; category II: 52.3% [41.7–62.0], 12.3 kcal/kg day [9.4–15.4]; category III: 100.0% [88.4–107.2], 23.5 kcal/kg day [19.5–27.8]). There was a strong correlation (r = 0.89) between the daily amount of calories and of protein/amino acids provided during artificial nutrition.

3.3. Association of nutrition with the 30-day risk of dying

The associations of the variables in the confounder model with short-term survival are presented in Fig. S1–S3, and in Table S2 of the Supplementary Appendix. There was no evidence that the number of nutrition days with parenteral nutrition or with oral intake (during the first three nutrition days) was connected with the risk of dying (HR 1.03, 95% CI 0.98 to 1.09 and 0.87, 95% CI 0.73 to 1.03, respectively). We could, however, identify time-varying associations between daily energy intake and outcome. Figures 1 and 2 show the results of the six comparisons of the five, different hypothetical nutrition support plans.

3.3.1. Comparison of a mildly hypocaloric with a severely hypocaloric nutrition (Fig. 1, and Table S3 of the Supplementary Appendix)

We compared three different hypothetical nutrition support plans: a complete (nutrition days 1 to day 11), severely hypocaloric nutrition support plan (plan #1, daily feeding of calories in the order of category I, about 1 kcal/kg); a delayed mildly hypocaloric nutrition support plan (plan #2, daily feeding of calories in the order of category I on nutrition days 1–4, and of category II on nutrition days 5–11); an early, mildly hypocaloric nutrition support plan (plan #3, daily feeding of calories in the order of category II, about 12 kcal/kg).

The key finding was that, when compared with the complete, severely hypocaloric nutrition support plan #1, the early, mildly hypocaloric nutrition support plan #3 was strongly associated with a better short term outcome (Fig. 1a). The hazard ratio began decreasing below 1 (favoring more nutrition) immediately after the 4 day lag-time reaching full effect by day 14 and remaining at 0.59 (95% CI, 0.37 to 0.93) thereafter (see Table S3 of the Supplementary Appendix).

Figure 1b depicts the daily hazard ratio comparing the support plan that starts severely hypocaloric for the first 5 days but thereafter remains mildly hypocaloric (plan # 2) to the nutrition support plan that remained severely hypocaloric (plan #1). In this comparison, a mortality benefit for plan #2 begins to appear after day 9 hitting a minimal hazard ratio by day 14 (HR = 0.70, 95% CI, 0.51 to 0.96) and remaining fairly constant thereafter (see Table S3 of the Supplementary Appendix).

Compared to nutrition support plan #2, early, mildly hypocaloric nutrition support plan #3 provided some evidence in favor of plan #3, albeit the null effect (HR = 1) was contained within the CIs over the entire follow-up period (Fig. 1c and Table S3, minimal hazard ratio 0.83, 95% CI 0.68 to 1.01).

3.3.2. Comparison of a near target caloric supply with hypocaloric nutrition (Fig. 2, and Table S4 of the Supplementary Appendix)

Comparisons of hypothetical nutrition support plans additionally included those which were either partially close to the caloric

Table 1
Definition of hypothetical nutrition support plans; categories I–III of daily nutritional intake (oral + artificial) correspond to different fractions of the caloric target calculated at ICU admission to guide artificial nutrition; (I: <30% of target + no oral intake; II: 30–70% of target, or <30% of target + additional oral intake; III: >70% of target ± additional oral intake, or 30–70% of target + additional oral intake).

	Plan	Definition
#1	Complete, severely hypocaloric nutrition support plan	Feeding of calories in the order of category I on nutrition days 1–11
#2	Delayed, mildly hypocaloric nutrition support plan	Feeding of calories in the order of category I on nutrition days 1–4, and of category II on nutrition days 5–11
#3	Early, mildly hypocaloric nutrition support plan	Feeding of calories in the order of category II on nutrition days 1–11
#4	Delayed, near target nutrition support plan	Feeding of calories in the order of category II on nutrition days 1–4, and of category III on nutrition days 5–11
#5	Early, near target nutrition support plan	Feeding of calories in the order of category III on nutrition days 1–11

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

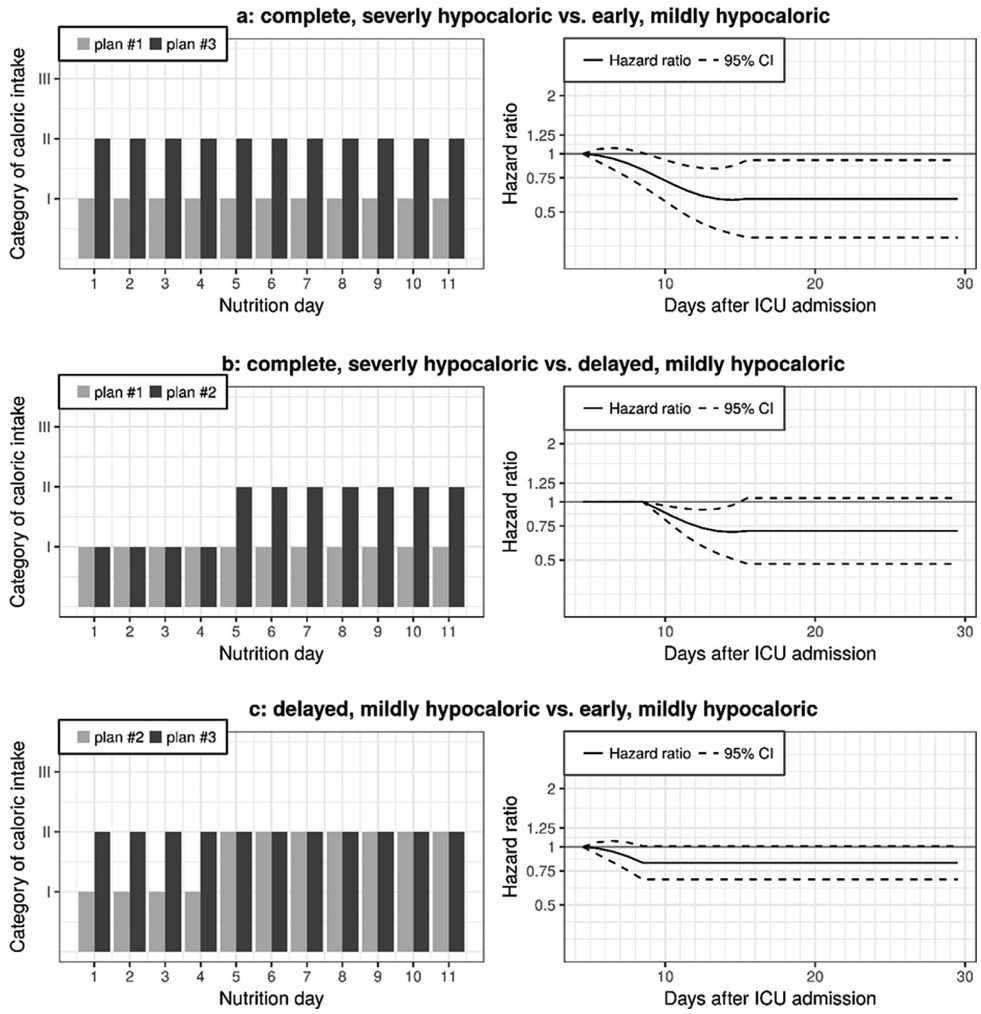


Fig. 1. Comparison of a mildly hypocaloric (category II) with a severely hypocaloric (category I) nutrition. Graphs shown on the left: design of plan comparisons analyzing different hypothetical nutrition support plans (Table 1). Categories of calorie intake were: C I, <30% of target for artificial nutrition (+no oral intake), about 1 kcal/kg day; C II, 30–70% of target, or <30% of target + additional oral intake, about 12 kcal/kg day. Graphs shown on the right: corresponding time-varying associations of different hypothetical nutrition support plans with the risk of dying. Solid lines indicate hazard ratios (HR), hatched lines indicate corresponding 95% confidence intervals (CI) (HRs and CIs for specific time intervals after ICU admission are presented in Table S3 of the Supplementary Appendix). Reference plan is that which provides fewer calories (e.g., an HR (and 95% CI) < 1 would indicate that the risk associated with the plan providing more calories was smaller). Please note that HRs (and corresponding 95% CIs) must be 1 for the first three time intervals (due to the specification of the lag time), and also for time intervals, in which nutritional categories of both hypothetical plans are identical.

target (daily feeding of calories in the order of category II (about 12 kcal/kg) on nutrition days 1–4, and of category III (plan #4, about 24 kcal/kg day) on nutrition days 5–11), or completely close to the target (plan #5, daily feeding of calories in the order of category III on nutrition days 1–11). The main analysis revealed that the early, near target nutrition support plan #5 was strongly associated with a reduced short-term risk when compared with the complete, severely hypocaloric nutrition support plan #1 (Fig. 2a, minimal hazard ratio 0.50, 95% CI 0.35 to 0.69 after the second week after ICU admission, Table S4 of the Supplementary Appendix).

There was no evidence, however, that both the delayed and the early, near target nutrition support plans #4 and #5 were associated with a lower 30-day risk of dying when compared with the early, mildly hypocaloric nutrition support plan #3 (category II) (Fig. 2b and c).

The findings of the main analysis were qualitatively compatible with the results of the sensitivity analyses performed to test the robustness of results with regard to different assumptions about survival after hospital discharge, or amount of oral calories (results for specific plan comparisons are presented in Figs. S5 and S6 of the Supplementary Appendix).

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

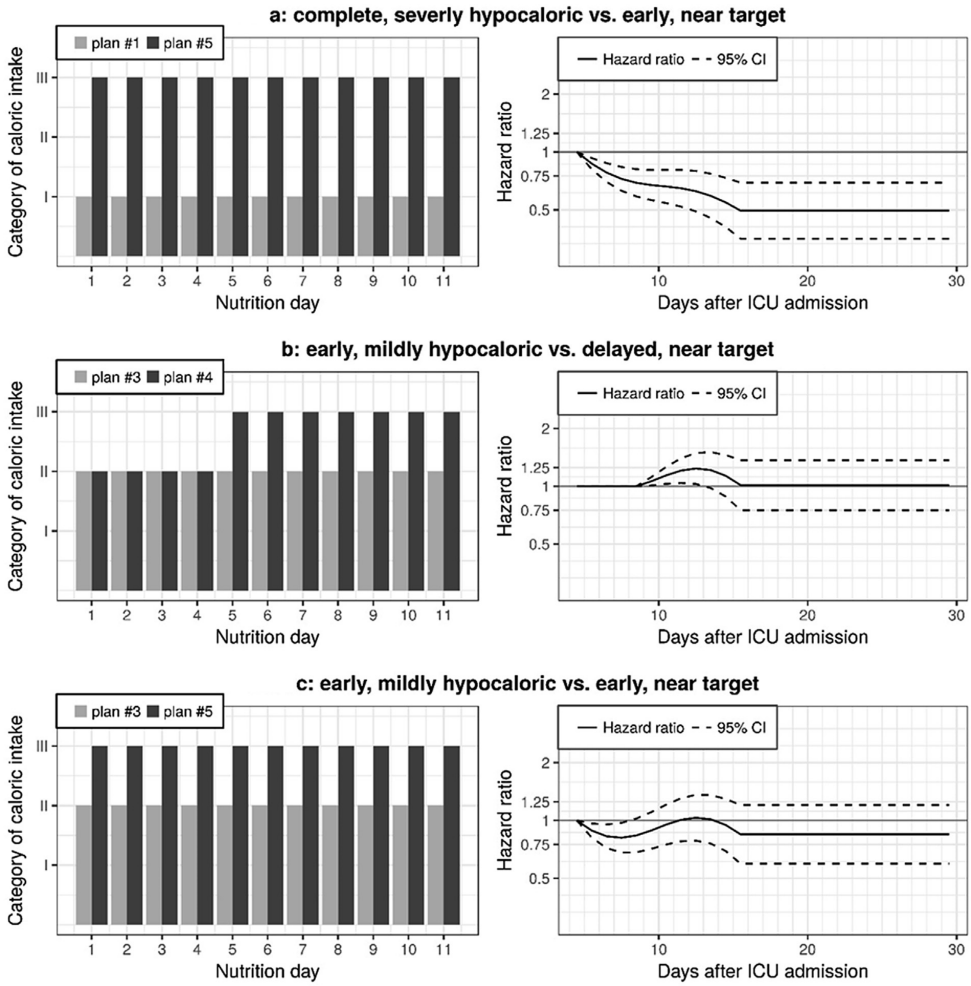


Fig. 2. Comparison of a near target (category III) with a hypocaloric (category I and II) nutrition. Graphs shown on the left: design of plan comparisons analyzing different hypothetical nutrition support plans (Table 1). Categories of calorie intake were: C I, <30% of target for artificial nutrition (+no oral intake) (about 1 kcal/kg day); C II, 30–70% of target, or <30% of target + additional oral intake (about 12 kcal/kg day); CIII >70% of target ± additional oral intake, or 30–70% of target + additional oral intake (about 24 kcal/kg day). Graphs shown on the right: corresponding time-varying associations of different hypothetical nutrition support plans with the risk of dying. Solid lines indicate hazard ratios (HR), hatched lines indicate corresponding 95% confidence intervals (CI) (HRs and CIs for specific time intervals after ICU admission are presented in Table S4 of the Supplementary Appendix). Reference plan is the one which provides fewer calories (e.g., an HR (and 95% CI) < 1 would indicate that the risk associated with the plan providing more calories was smaller). Please note that HRs (and corresponding 95% CIs) must be 1 for the first three time intervals (according to the specification of lag time), and for time intervals, in which nutritional categories of both hypothetical plans are identical.

3.3.3. Association of nutritional intake with the risk of dying in subgroups

Due to comparatively low patient numbers, 95% confidence intervals were wide emphasizing the clearly hypothesis-generating character of the results. In patients with Apache II score values > 25 (n = 3200) associations between different amounts of calories and mortality generally appeared to be somewhat weaker (e.g., hazard ratios closer to 1) than in the whole cohort (Fig. S7). Associations appeared to differ between medical (n = 6181) and surgical patients (n = 3480) (Figs. S8 and S9), and between patients being overweight (BMI ≥ 25, n = 5332) or not (n = 4329) (Figs. S10 and S11). Results of medical or overweight patients were qualitatively comparable to that found in the whole cohort, whereas surgical patients appeared to have been largely insensitive to variations of daily caloric intake. In patients with a BMI < 25 (including underweight patients), an early, near target nutrition support plan #5 (Category III) was associated with a reduced short-term risk when compared with an early, mildly hypocaloric nutrition support plan #3 (category II), whereas no such association could be found when comparing an early, mildly hypocaloric nutrition support plan with a complete, severely hypocaloric nutrition support plan #1 (category I).

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

Table 2
Demographic and clinical characteristics (categorical variables) of the patients (MV, mechanical ventilation; OI, oral intake; EN, enteral nutrition; PN, parenteral nutrition, PF, propofol; the duration/number of days were calculated up to nutrition day 3).

		All		Outcome	
		n	Percent	Surviving[a]	Surviving (%)
Year	2007	2122	22.0	1686	79.5
	2008	2056	21.3	1606	78.1
	2009	2308	23.9	1860	80.6
	2011	3175	32.9	2535	79.8
Gender	Female	3847	39.8	3033	78.8
	Male	5814	60.2	4654	80.0
Duration of MV	<1 h	48	0.5	44	91.7
	1–24 h	232	2.4	213	91.8
	25–48 h	474	4.9	415	87.6
	>48 h	8907	92.2	7015	78.8
Number of days with OI	0	9179	95.0	7274	79.2
	1	308	3.2	261	84.7
	2	134	1.4	113	84.3
	3	40	0.4	39	97.5
Number of days with PF therapy	0	5807	60.1	4495	77.4
	1	989	10.2	801	81.0
	2	998	10.3	834	83.6
	3	1867	19.3	1557	83.4
Duration of PN	<1 h	7860	81.4	6255	79.6
	1–24 h	302	3.1	238	78.8
	25–48 h	444	4.6	356	80.2
	>48 h	1055	10.9	838	83.7
Duration of EN	<1 h	1914	19.8	1489	77.8
	1–24 h	924	9.6	742	80.3
	25–48 h	1933	20.0	1519	78.6
	>48 h	4890	50.6	3937	80.5
Admission category	Surgical/Elective	1073	11.1	913	85.1
	Medical	6181	64.0	4748	76.8
	Surgical/Emergency	2407	24.9	2026	84.2
Admission diagnosis	Gastrointestinal	1465	15.2	1183	80.8
	Cardio-Vascular	1440	14.9	1080	75.0
	Other	476	4.9	384	80.7
	Metabolic	199	2.1	182	91.5
	Neurologic	1269	13.1	1029	81.1
	Orthopedic/Trauma	1117	11.6	995	89.1
	Renal	104	1.1	77	74.0
	Respiratory	2618	27.1	2051	78.3
	Sepsis	974	10.1	706	72.5

[a] Surviving patients were those either surviving until discharge from the hospital, or until day 30 while still being hospitalized.

Table 3
Demographic and clinical characteristics (quantitative variables) of the patients.

	Age (years)	Apache II Score	Body mass index (kg/m^2)
Min.	18.0	0.0	13.1
1st Qu.	48.0	17.0	22.6
Median	62.0	22.0	25.7
Mean	59.5	22.5	27.3
3rd Qu.	73.0	28.0	30.1
Max.	100.0	71.0	104.8

4. Discussion

The results of our study suggest that in a sample of heterogeneous critically ill patients remaining in the ICU for at least 96 h, provision of only minimal amounts of calories (category I, <30% of target, about 1 kcal/kg day) throughout the first 11 nutrition days is associated with an increased risk of dying during the first 30 days after ICU admission. Compared to such an extremely low calorie intake, a mildly hypocaloric nutrition (category II, 30–70% of target, about 12 kcal/kg day) was strongly associated with a mortality risk reduction.

The precise minimum caloric requirement, however, is disputable. There were associations between a gradually increasing number of nutrition days with a mildly hypocaloric nutrition and a better outcome. This leaves open the possibility of an unchanged 30-day mortality risk if minimal amounts of calories (<30% of target) were provided on only one or two nutrition days.

A second key finding of our study was that, compared to a baseline supply of mildly hypocaloric amounts of calories (category II), a delayed (on days 5–11) or an early (on days 1–11) supply of calories in the range of category III (>70% of target, about 24 kcal/kg day) was not associated with further short-term mortality benefits.

It is currently highly controversial whether a near target caloric supply is beneficial or not. Strong associations between a near target calorie intake and a lower mortality have only been identified by some observational studies [5,15,16]. These findings were not universal, however, and other observational studies found beneficial associations only in subjects who were severely overweight (BMI > 35), malnourished (BMI < 25) [10], or at a high nutritional risk [17], or even presented contradictory results (an association of a near target intake with a higher mortality) [18–22]. Interpretation of these results is extremely difficult since none of these studies specifically considered reverse causation phenomena, or adjusted for duration, type and daily variation of energy intake.

Our study attempts to address several analytical limitations (such as indication bias) inherent in the design of older observational studies, specifically by adjusting for multiple sources of confounding and taking into account the time-dependent nature of

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

caloric intake. Instead of using average calorie intake during the observation period, associations with outcome were based on daily amounts of calorie intake. Additionally, we adjusted associations of calorie intake with outcome to the use of additional parenteral or oral nutrition.

To corroborate our analytical strategy, we performed a sensitivity analysis testing the importance of additional oral calorie intake on a specific nutrition day for outcome. Since we found that oral intake was unimportant for outcome (Fig. S8), we can be reasonably certain that our analytical strategy largely eliminated bias resulting from variations in the route by which calories had been supplied. Furthermore, these results suggest that further investigations of similar data could ignore oral intake and focus on the continuous effect of artificial calorie intake.

Based on that novel analytical strategy, our study yielded results, which are largely in line with the findings of the randomized studies. Randomized studies were analyzed recently by nine meta-analyses comparing a) a severely (<30% of target, <8 kcal/kg day) with a mildly hypocaloric energy intake (30–60% of target, ≈8–16 kcal/kg day) [23], b) a mildly hypocaloric energy intake with the provision of slightly hypocaloric amounts of calories (70–80% of target, ≈16–21 kcal/kg day) [24–30] or c) a mildly hypocaloric with an isocaloric energy intake (100% of target, ≈24 kcal/kg day) [9]. Corresponding to our findings, two meta-analyses showed that providing only minimal amounts of calories (e.g., <33% of target) is detrimental and that giving 30–60% of target calories will decrease mortality [26,28]. Compared to a mildly hypocaloric nutrition, however, provision of more calories either did not affect mortality [23–25,27,29,30] or even worsened it [26,28].

Altogether, it is likely that in the early phase after an insult (day 1–11) daily provision of 30–70% of the daily caloric target (according to our study, about 10–15 kcal/kg day) would be sufficient to minimize 30-day mortality, and that provision of more calories would not convey a further risk reduction, but may even be detrimental. Our subgroup analyses suggest that interactions between selected covariates and nutrition might be important for the absence or presence of effects. Thus, surgical patients and patients with a high risk (Apache II Score values > 25) were generally less responsive to different amounts of calories than the whole cohort, whereas body weight might modify nutrition effects in a way that overweight patients require less and underweight patients more calories.

4.1. Limitations and strengths

Our study has certain limitations. Calorie intake in our study was rarely guided by indirect calorimetry. Equations to estimate energy expenditure may overestimate the energy needs in up to 70% of patients [31].

We do not know whether associations between caloric supply and short-term mortality would also hold for morbidity. According to the meta-analyses, however, effects on morbidity may also parallel those on mortality. Thus, a severely hypocaloric nutrition was found to increase the infection rate [23], and – compared to a mildly hypocaloric nutrition – provision of more calories did not alter morbidity (rates of new infections, duration of mechanical ventilation, length of stay) in five meta-analyses [9,25,26,30]. Three meta-analyses, however, reported an increased morbidity [24,27,29], thereby largely excluding the possibility that – in unselected critically ill patients - the provision of isocaloric or nearly isocaloric amounts of calories might be beneficial during the acute phase of the disease.

We also could not separately analyze the contribution of energy versus protein intake to mortality. Finally, due to the observational nature of the study, we cannot entirely exclude the possibility that our results are – at least partially – affected by unmeasured confounding or residual indication bias. Indication bias may originate from the phenomenon that a better health is associated with a gastrointestinal tract working more efficiently thereby allowing the supply of more enteral/oral calories. Thereby, a higher calorie intake may be associated with a better outcome without being the true cause. Similarly, bias could result from possible changes of calorie intake just prior to death. The 4-day lag between nutrition and its' effect on the outcome, as used in our analyses, should avoid the most severe effects from such indication bias, as, for example, hypocaloric nutrition just prior to death would not enter the analysis.

The strengths of our study are its size, allowing an in-depth examination of the association between caloric supply and short-term survival, and its statistical strategies. We used novel statistical techniques designed to estimate the time-dependent association of caloric intake and outcome. By including a wide variety of diagnoses and diseases this study also overcomes the criticisms expressed by the opponents of randomized studies. Furthermore, for the first time we could calculate the effect of an extreme scenario, namely the provision of extremely low amounts of calories during the first 11 days after ICU admission. The large number of events and of participating ICUs from many countries may help support the generalizability of the findings.

Funding

Fabian Scheipl was supported by the German Research Foundation through the Emmy Noether Programme (GR 3793/1-1 to Sonja Greven).

Authors' contributions

WH, FS, AB and HK developed the statistical analysis strategy. AB and FS performed the statistical modeling and inference. AB and AD were responsible for data management and created descriptive statistics and visualizations. WH and DK did the literature search and drafted the manuscript. All authors contributed to interpretation of the data and critical revision of the manuscript.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgement

The authors thank Daren Heyland, Clinical Evaluation Research Unit, Kingston General Hospital, Kingston, Ontario, Canada, for kindly providing data from the international “nutrition practice in ICUs” registry.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.clnu.2018.04.005.

References

[1] Casaer MP, Mesotten D, Hermans G, Wouters PJ, Schetz M, Meyfroidt G, et al. Early versus late parenteral nutrition in critically ill adults. N Engl J Med 2011 Aug 11;365(6):506–17.
[2] National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network, Rice TW, Wheeler AP, Thompson BT, Steingrub J, Hite RD, Moss M, et al. Initial trophic vs full enteral feeding in

patients with acute lung injury: the EDEN randomized trial. J Am Med Assoc 2012 Feb 22;307(8):795–803.
[3] McClave SA, Heyland DK, Martindale RG. Adding supplemental parenteral nutrition to hypocaloric enteral nutrition: lessons learned from the Casaer Van den Berghe study. J Parenter Enteral Nutr 2012 Jan;36(1):15–7.
[4] Schetz M, Casaer MP, Van den Berghe G. Does artificial nutrition improve outcome of critical illness? Crit Care 2013 Feb 1;17(1):302.
[5] Heyland DK, Cahill N, Day AG. Optimal amount of calories for critically ill patients: depends on how you slice the cake! Crit Care Med 2011 Dec;39(12):2619–26.
[6] Singer P, Hiesmayr M, Biolo G, Felbinger TW, Berger MM, Goeters C, et al. Pragmatic approach to nutrition in the ICU: expert opinion regarding which calorie protein target. Clin Nutr 2014 Apr;33(2):246–51.
[7] McClave SA, Taylor BE, Martindale RG, Warren MM, Johnson DR, Braunschweig C, et al., Society of Critical Care Medicine, American Society for Parenteral and Enteral Nutrition. Guidelines for the provision and assessment of nutrition support therapy in the adult critically ill patient: Society of Critical Care Medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.). J Parenter Enteral Nutr 2016 Feb;40(2):159–211.
[8] Canadian critical care nutrition clinical practice guidelines. 2015. http://www.criticalcarenutrition.com/cpgs.
[9] Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Crit Care Med 2017 Mar;45(3):486–552.
[10] Alberda C, Gramlich L, Jones N, Jeejeebhoy K, Day AG, Dhaliwal R, et al. The relationship between nutritional intake and clinical outcomes in critically ill patients: results of an international multicenter observational study. Intensive Care Med 2009 Oct;35(10):1728–37.
[11] Van den Berghe G, Wilmer A, Hermans G, Meersseman W, Wouters PJ, Milants I, et al. Intensive insulin therapy in the medical ICU. N Engl J Med 2006 Feb 2;354(5):449–61.
[12] Bender A, Scheipl F, Hartl W, Day AG, Küchenhoff H. Penalized estimation of complex, non-linear exposure-lag-response associations. Biostatistics 2018 Feb 12. https://doi.org/10.1093/biostatistics/kxy003 [Epub ahead of print].
[13] Friedman M. Piecewise exponential models for survival data with covariates. Ann Stat 1982;10(1):101–13.
[14] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. http://www.R-project.org/.
[15] Elke G, Wang M, Weiler N, Day AG, Heyland DK. Close to recommended caloric and protein intake by enteral nutrition is associated with better clinical outcome of critically ill septic patients: secondary analysis of a large international nutrition database. Crit Care 2014 Feb 10;18(1):R29.
[16] Weijs PJ, Stapel SN, de Groot SD, Driessen RH, de Jong E, Girbes AR, et al. Optimal protein and energy nutrition decreases mortality in mechanically ventilated, critically ill patients: a prospective observational cohort study. J Parenter Enteral Nutr 2012 Jan;36(1):60–8.
[17] Compher C, Chittams J, Sammarco T, Nicolo M, Heyland DK, et al. Greater protein and energy intake may be associated with improved mortality in higher risk critically ill patients: a multicenter, multinational observational study. Crit Care Med 2017 Feb;45(2):156–63.
[18] Kutsogiannis J, Alberda C, Gramlich L, Cahill NE, Wang M, Day AG, et al. Early use of supplemental parenteral nutrition in critically ill patients: results of an international multicenter observational study. Crit Care Med 2011 Dec;39(12):2691–9.
[19] Arabi YM, Haddad SH, Tamim HM, Rishu AH, Sakkijha MH, Kahoul SH, et al. Near-target calorie intake in critically ill medical-surgical patients is associated with adverse outcomes. J Parenter Enteral Nutr 2010 May-Jun;34(3):280–8.
[20] Elke G, Kuhnt E, Ragaller M, Schädler D, Frerichs I, Brunkhorst FM, et al. Enteral nutrition is associated with improved outcome in patients with severe sepsis. A secondary analysis of the VISEP trial. Med Klin Intensivmed Notfmed 2013 Apr;108(3):223–33.
[21] Zusman O, Theilla M, Cohen J, Kagan I, Bendavid I, Singer P, et al. Resting energy expenditure, calorie and protein consumption in critically ill patients: a retrospective cohort study. Crit Care 2016 Nov 10;20(1):367.
[22] Crosara IC, Mélot C, Preiser JC. A J-shaped relationship between caloric intake and survival in critically ill patients. Ann Intensive Care 2015 Dec;5(1):37.
[23] Reintam Blaser A, Starkopf J, Alhazzani W, Berger MM, Casaer MP, Deane AM, et al. ESICM Working Group on Gastrointestinal Function. Early enteral nutrition in critically ill patients: ESICM clinical practice guidelines. Intensive Care Med 2017 Mar;43(3):380–98.
[24] Al-Dorzi HM, Albarrak A, Ferwana M, Murad MH, Arabi YM. Lower versus higher dose of enteral caloric intake in adult critically ill patients: a systematic review and meta-analysis. Crit Care 2016;20:358.
[25] Marik PE, Hooper MH. Normocaloric versus hypocaloric feeding on the outcomes of ICU patients: a systematic review and meta-analysis. Intensive Care Med 2016 Mar;42(3):316–23.
[26] Choi EY, Park DA, Park J. Calorie intake of enteral nutrition and clinical outcomes in acutely critically ill patients: a meta-analysis of randomized controlled trials. J Parenter Enteral Nutr 2015 Mar;39(3):291–300.
[27] Parikh HG, Miller A, Chapman M, Moran JL, Peake SL, et al. Calorie delivery and clinical outcomes in the critically ill: a systematic review and meta-analysis. Crit Care Resuscitation 2016 Mar;18(1):17–24.
[28] Tian F, Wang X, Gao X, Wan X, Wu C, Zhang L, et al. Effect of initial calorie intake via enteral nutrition in critical illness: a meta-analysis of randomised controlled trials. Crit Care 2015 Apr 20;19:180.
[29] Chelkeba L, Mojtahedzadeh M, Mekonnen Z. Effect of calories delivered on clinical outcomes in critically ill patients: systemic review and meta-analysis. Indian J Crit Care Med 2017 Jun;21(6):376–90.
[30] Ridley EJ, Davies AR, Hodgson CL, Deane A, Bailey M, Cooper DJ, et al. Delivery of full predicted energy from nutrition and the effect on mortality in critically ill adults: a systematic review and meta-analysis of randomised controlled trials. Clin Nutr 2017 Oct 9. pii: S0261-5614(17) 31358-4.
[31] Pichard C, Oshima T, Berger MM. Energy deficit is clinically relevant for critically ill patients: yes. Intensive Care Med 2015 Feb;41(2):335–8.

Please cite this article in press as: Hartl WH, et al., Calorie intake and short-term survival of critically ill patients, Clinical Nutrition (2018), https://doi.org/10.1016/j.clnu.2018.04.005

Part III

Software

Chapter 5

Piece-wise exponential Additive Mixed Modeling Tools

Chapter 5 introduces the **R** package **pammtools**, which provides functions that facilitate the data transformation, visualization and interpretation of Piece-wise exponential Additive Mixed Models of different complexity, including time-varying effects and cumulative effects.

Contributing article:

Bender, A. and Scheipl, F. (2018b). pammtools: Piece-wise exponential additive mixed modeling tools. *arXiv:1806.01042 [stat]*

Author contributions:

Andreas Bender created the package and was responsible for the design and implementation. Fabian Scheipl gave valuable input on design choices and contributed to the package. Andreas Bender prepared the manuscript including examples and visualizations. Fabian Scheipl contributed to the final draft and proofread the manuscript.

Supplementary material available at:

`https://adibender.github.io/pammtools/`

pammtools: Piece-wise exponential Additive Mixed Modeling tools

Andreas Bender
LMU Munich
Universitätsklinikum Regensburg

Fabian Scheipl
LMU Munich

Abstract

This article introduces the **pammtools** package, which facilitates data transformation, estimation and interpretation of Piece-wise exponential Additive Mixed Models. A special focus is on time-varying effects and cumulative effects of time-dependent covariates, where multiple past observations of a covariate can cumulatively affect the hazard, possibly weighted by a non-linear function. The package provides functions for convenient simulation and visualization of such effects as well as a robust and versatile function to transform time-to-event data from standard formats to a format suitable for their estimation. The models can be represented as Generalized Additive Mixed Models and estimated using the **R** package **mgcv**. Many examples on real and simulated data as well as the respective **R** code are provided throughout the article.

Keywords: survival analysis, time-varying effects, time-dependent covariates, cumulative effects, distributed lags, exposure-lag-response associations, functional data analysis, generalized additive mixed models.

1. Introduction

This article introduces the **pammtools** package (https://adibender.github.io/pammtools/), which provides functions to facilitate the estimation and interpretation of a class of models for time-to-event data analysis, which we call *P*iece-wise exponential *A*dditive *M*ixed *M*odels (PAMMs; Bender, Groll, and Scheipl 2018a). PAMMs are a semi-parametric extension of the Piece-wise Exponential Model (PEM) (Friedman 1982) that allow for penalized estimation of very flexible survival models with (time-varying, non-linear) covariate effects, random effects and cumulative effects of time-varying covariates, also known as distributed lags and exposure-lag-response associations (Gasparrini 2014). In short, PAMMs directly transfer the flexibility and performance available in current implementations of generalized additive regression models (GAMs) to time-to-event models.

Using PAMMs for time-to-event data analysis involves three main steps

1. **Data pre-processing**: This can be more or less involved, depending on the type of effects one wants to estimate (especially when the goal is to estimate cumulative effects) and depending on the type of software/package one wants to use for estimation (cf. Section 3).
2. **Estimation**: This step is currently performed outside of **pammtools**. In this article we use **mgcv** (Wood 2011) for estimation but any other package that implements

GAMMs or variants thereof can also be used, e.g., model-based boosting via **mboost** (Hothorn and Bühlmann 2006; Hofner, Mayr, Robinzonov, and Schmid 2012). Most post-processing and visualization functions in **pammtools** are customized to work with `mgcv::gam` objects, however.

3. **Model post-processing**: This includes calculation of estimated hazard rates, cumulative hazards and survival probabilities, which all need to take into account the specific data structure of PAMMs, as well as model/effect visualization, which can also become relatively complex, again, especially in the case of cumulative effects.

In the following, Section 2 briefly describes the piece-wise exponential additive mixed model and introduces the notation used throughout this article. Section 3 demonstrates the data transformations necessary to fit PAMMs in different scenarios, i.e., for data with and without time-dependent covariates (TDCs) and depending on the type of effects to be estimated. In Section 4, we discuss some application examples on real and simulated data to illustrate the estimation, visualization and interpretation of the different effect types in (1), facilitated by convenience functions provided in **pammtools**. Throughout, the results obtained by PAMMs are compared to estimates obtained from other established models when applicable.

For the code examples, the following packages will be used:

```
devtools::install_github("adibender/pammtools")
library(dplyr); library(tidyr); library(purrr); library(ggplot2)
library(survival); library(mgcv); library(pammtools)
```

2. Piece-wise Exponential Additive Mixed Models

In this article, we consider models for time-to-event analysis with hazard rates given by (1) and in the log-linear form by (2). Note that in (2) the log-baseline hazard was split in two terms such that $\log(\lambda_0(t)) = \beta_0 + f_0(t)$.

$$\lambda_i(t; \mathbf{x}_i, \mathcal{Z}_i, \ell_i) = \lambda_0(t) \exp\left[\sum_{p=1}^{P} f_p(x_{i,p}, t) + \sum_{m=1}^{M} g(\mathbf{z}_{i,m}, t) + b_{\ell_i}\right] \qquad (1)$$

$$i = 1, \ldots, n$$

$$\log(\lambda_i(t; \mathbf{x}_i, \mathcal{Z}_i, \ell_i)) = \beta_0 + f_0(t) + \sum_{p=1}^{P} f_p(x_{i,p}, t) + \sum_{m=1}^{M} g(\mathbf{z}_{i,m}, t) + b_{\ell_i} \qquad (2)$$

Let T_i and C_i, the true event and censoring times of subject i, respectively. Then, $\langle t_i, \delta_i \rangle$ is the observed event time tuple for subject i with event time $t_i = \min(T_i, C_i)$ and status indicator $\delta_i = I(T_i \leq C_i)$, $\mathbf{x}_i$ is the vector of time-constant covariates $x_{i,p}, p = 1, \ldots, P$ and $\mathcal{Z}_i = \{\mathbf{z}_{i,m} : m = 1 \ldots, M\}$ is the set of M time-dependent covariate vectors (exposure histories) $\mathbf{z}_{i,m} = \{z_{i,m}(t_{z_m,q_m}) : q_m = 1, \ldots, Q_m\}$, where t_{z_m} the (exposure) time points at which covariate z_m was observed. It is important to stress the difference between t, which denotes the time scale on which the event times are observed and t_z, which denotes the time scale

on which time-dependent covariate z is observed. The two scales t and t_z do not need to be identical or even overlap, nor do they have to be measured in the same units (see Section 3.3 and 4.3 for examples).

In the following paragraphs, we briefly describe the individual components in (2). A tutorial style exposition of the model without the $g(\mathbf{z}, t)$ terms is given in Bender *et al.* (2018a) and a very general framework for models with cumulative effects $g(\mathbf{z}, t)$ is described and evaluated in Bender, Scheipl, Hartl, Day, and Küchenhoff (2018b).

The terms $f_p(x_{i,p}, t)$ denote time-varying effects (TVEs) of time-constant covariates $\mathbf{x}_{\cdot,p}$, and our notation subsumes the entire range of effects of this kind, i.e., from time-constant linear effects all the way to non-linear and non-linearly time varying effect surfaces and everything in between. A selection of possible TVEs along with their specification for estimation with `mgcv::gam` are summarized in Table 1. Note that models with multiple time-varying effects may need to impose additional identifiability constraints (Wood 2017, Ch. 5.6.3), see `?mgcv::ti` and the examples in Section 4.2. Also note that (non-linear, non-linearly time-varying) interaction effects of multiple covariates can be specified and estimated in the same way in this framework.

The terms $g(\mathbf{z}_{i,m}, t)$ are potentially (non-linearly) time-varying, potentially cumulative effects of time dependent covariates $\mathbf{z}_{\cdot,m}$. Such terms are discussed in more detail in Sections 3.2 (data transformation) and 4.3 (modeling).

The term b_{ℓ_i} denotes random effects (a log-normal frailty) associated with group $\ell = 1, \ldots, L$ to which subject i belongs. Extensions to more complex random effect structures for nested or crossed groups or spatial effects are possible within the presented framework as well (e.g. Wood and Scheipl 2017). For an example of a random effect model estimated via PAMMs see the frailty vignette. In the following, we omit the random effect term to focus on time-varying and cumulative effects rather than hierarchical models.

Table 1: Selection of possible $f(x_{i,p}, t)$ effect specifications in PAMMs, including the **R** code when fitted using `mgcv::gam`. Here `x` denotes any covariate of interest in the data set and `t` a representation of time in each interval. This could be for example the interval end-points $t_j := \kappa_j$ or interval mid-points $t_j := (\kappa_{j-1} + (\kappa_j - \kappa_{j-1})/2)$.

$f(x_{i,p}, t) =$	Description	Specification in `mgcv::gam`
$\beta_p x_{i,p} \cdot x_{i,p}$	Linear, time-constant effect	`... + x + ...`
$f_p(x_{i,p})$	Smooth nonlinear, time-constant effect	`... + s(x) + ...`
$\beta_p x_{i,p} + \beta_{p:t} \cdot x_{i,p} \cdot t$	Linear, linearly time-varying effect	`... + x + x:t ...`
$f_p(x_{i,p}) \cdot t$	Smooth, linearly time-varying effect	`... + s(x, by=t) + ...`
$x_{i,p} \cdot f_p(t)$	Linear, smoothly time-varying effect	`... + s(t,by=x) + ...`
$f_p(x_{i,p}, t)$	Smooth, smoothly time-varying effect	`... + te(x,t) + ...`

To estimate model (1) using PAMMs, the time under risk is divided into J intervals with interval cut points $\kappa_0 < \ldots < \kappa_J$ that define intervals $(\kappa_{j-1}, \kappa_j], j = 1 \ldots, J$. The smooth hazard $\lambda(t)$ is approximated by piece-wise constant hazards $\lambda(t) = \lambda(t_j) \forall t \in (\kappa_{j-1}, \kappa_j]$ where $t_j \in (\kappa_{j-1}, \kappa_j]$ denotes any fixed timepoint in the j-th interval, (typically $t_j := \kappa_j$), such that

$$\log(\lambda_i(t; \mathbf{x}_i, \mathcal{Z}_i)) \approx \lambda_{ij} := \log(\lambda_i(t_j; \mathbf{x}_i, \mathcal{Z}_i)) \;\; \forall t \in (\kappa_{j-1}, \kappa_j], i = 1, \ldots, n \tag{3}$$

$$\approx \beta_0 + f_0(t_j) + \sum_{p=1}^{P} f_p(x_{i,p}, t_j) + \sum_{m=1}^{M} g(\mathbf{z}_{i,m}, t_j) \tag{4}$$

Piecewise constant hazard rates imply a piecewise exponential distribution of event times, thus: PEM and PAMM, but note that any shape of the conditional hazard rate can be approximated arbitrarily closely by a sufficiently dense step function.

In the classical PEM, the number of intervals J as well as the positioning of cut points κ_j are important parameters that affect the quality of the approximation (Demarqui, Loschi, and Colosimo 2008). This is less important for PAMMs as long as J is not to small and κ_j are sufficiently dense in areas where $\lambda(t; \mathbf{x}, \mathcal{Z})$ varies more quickly. In agreement with Whitehead (1980), we recommend to use the unique observed event and/or censoring times as cut-points, which automatically leads to improved approximation with increasing n and high κ_j density in the relevant parts of the follow-up. The default in **pammtools** is to use the uniquely observed event times. For large data sets, an exception to this rule might be preferable if computational resources are insufficient for the resulting data size. GAMMs for big data (cf. Wood, Goude, and Shaw (2015) and `?mgcv::bam`) are very useful in this context to reduce both memory load and computation time.

Regardless of the splitting scheme, once the interval split points κ_j are chosen, the data has to be transformed to what we call the piece-wise exponential data (PED) format (cf. Bender *et al.* (2018a) and the data-transformation vignette) with

- interval specific event indicators $\delta_{ij} = \begin{cases} 1, & \text{if } t_i \in (\kappa_{j-1}, \kappa_j] \wedge \delta_i = 1 \\ 0, & \text{else} \end{cases}$, and
- offsets $o_{ij} = \log(t_{ij})$, where $t_{ij} = \min(\kappa_j - \kappa_{j-1}, t_i - \kappa_{j-1})$

After this data transformation, the model can be estimated using Poisson regression with offsets o_{ij} under the working assumption $\delta_{ij} \overset{i.i.d.}{\sim} Po(\mu_{ij})$ with $\mu_{ij} = \lambda_{ij} t_{ij}$ and λ_{ij} as defined in (3), even though the working assumption of independent δ_{ij} is clearly violated (see Holford (1980); Whitehead (1980); Laird and Olivier (1981); Friedman (1982) for the original justification of the PEM and Cai, Hyndman, and Wand (2002); Kauermann (2005); Argyropoulos and Unruh (2015); Bender *et al.* (2018b) for penalized and mixed model based approaches).

3. Data pre-processing

Using pseudo-Poisson responses for time-to-event analysis requires a specific augmented data format called piece-wise exponential data (PED) in the following. **pammtools** provides convenience functions that perform this data augmentation to create the required additional covariates (e.g., $t_j := \kappa_j$, event indicators δ_{ij} and the offsets o_{ij}).

In the context of PAMMs, data transformation depends on the type of covariates that are present (time-constant (TCC) vs. time-dependent (TDC)) and the type of effects one wants to

estimate (time-constant or time-varying for TCCs and concurrent or cumulative for TDCs). In PAMMs, time-varying effects of TCCs are simply interactions of the covariates with (a function of) time. Therefore, no special treatment is required. Thus, we differentiate the following situations

- TCCs with potentially time-varying effects $f(t, x)$, see Section 3.1
- TDCs with concurrent (time-varying) effects $f(t)z(t)$, see Section 3.2
- TDCs with cumulative effects $\int_{\mathcal{T}(t)} h(t, t_z, z(t_z))\mathrm{d}t_z$, see Section 3.3

For all data transformations listed above, **pammtools** provides a single function `as_ped` (mnemonic: *as piece-wise exponential data*), with a formula based interface, which contains specials `concurrent` and/or `cumulative` in the presence of TDCs.

3.1. Time-constant covariates

In this section we illustrate the transformation of standard time-to-event data without TDCs to the PED format. All examples in this section will use the `tumor` data available in **pammtools**. The application of `as_ped` and its output are illustrated in **R**-chunk 1 for the first 2 rows for each category of the `sex` variable of the `tumor` data, using a rather crude 200-day partition of the follow up.

R-chunk 1

```
tumor_sub <- tumor %>% select(1:5) %>% group_by(sex) %>% slice(1:2)
tumor_sub

# A tibble: 4 x 5
# Groups:   sex [2]
   days status charlson_score   age sex
  <dbl>  <int>          <int> <int> <fct>
1 1192.      0              2    52 male
2   33.      1              2    57 male
3  579.      0              2    58 female
4  308.      1              2    74 female

ped <- tumor_sub %>%
  as_ped(Surv(days, status) ~., cut = seq(0,1000, by = 200)) %>%
  select(1:9)
ped

   id tstart tend   interval   offset ped_status charlson_score age    sex
1   1      0  200    (0,200] 5.298317          0              2  52   male
2   1    200  400  (200,400] 5.298317          0              2  52   male
3   1    400  600  (400,600] 5.298317          0              2  52   male
4   1    600  800  (600,800] 5.298317          0              2  52   male
5   1    800 1000 (800,1000] 5.298317          0              2  52   male
6   2      0  200    (0,200] 3.496508          1              2  57   male
7   3      0  200    (0,200] 5.298317          0              2  58 female
8   3    200  400  (200,400] 5.298317          0              2  58 female
9   3    400  600  (400,600] 5.187386          0              2  58 female
10  4      0  200    (0,200] 5.298317          0              2  74 female
11  4    200  400  (200,400] 4.682131          1              2  74 female
```

In the `as_ped` call in **R**-chunk 1

- the left hand side (LHS) of the `formula` specifies the event time and status information. Currently **pammtools** only supports right-censored data.
- the right hand side (RHS) of the `formula` specifies covariates that should be kept after data transformation. This can be useful when the data contains many variables but only a few will be used to estimate the hazard. As usual, a dot (`~.`) can be used to include all variables.
- the follow up is partitioned at the split points κ_j provided through the `cut` argument. The start (`tstart`) and stop (`tend`) times are created as well as an `interval` column.
- the δ_{ij}, which will serve as the outcome of the Poisson regression, are stored in the column `ped_status` and are 1 only in the interval in which the subject experienced an event (if uncensored), which is also the final interval for that subject.
- the offset variable is calculated, e.g., subject `id = 3` was censored at 579 days, therefore $o_{i=3,j=3} = \log(\min(579 - 400, 600 - 400)) = \log(179) = 5.187386$.
- subjects with event times $t_i > \kappa_J$ will be administratively censored at κ_J (see `id = 1`).

The output data has class `ped` and **pammtools** contains several `S3` methods that dispatch on `ped` objects. Examples are provided in Section 4, especially Section 4.4.

In **R**-chunk 2, `as_ped` is applied to all observations of the `tumor` data. As the `cut` argument is not explicitly specified, all unique t_i where $\delta_i = 1$ will be used as interval split points. The argument `max_time = 3034` indicates that the last interval should end at 3034 days, which means that all observations with $t_i > 3034$ will be considered censored at $\kappa_J = 3034$. This can be useful to limit the follow-up to a reasonable range with enough observations (i.e., events), which can make estimation of models faster and more robust, especially with respect to time dependent terms. Here, `max_time` was set to the last observed event time in order to facilitate comparisons to the Aalen model in Section 4.2.

R-chunk 2

```
ped_tumor <- tumor %>% as_ped(Surv(days, status)~., max_time =3034)
```

The data set `ped_tumor` will be used for illustration of the estimation and interpretation of time-constant effects and (non-linearly) time-varying effects in Sections 4.1 and 4.2, respectively.

3.2. Time-dependent covariates with concurrent effects

Transformation of data containing time-dependent covariates involves a little more work, as, usually, the interval split points κ_j are now the union of the user-specified split points and the time points at which (changes in) the time-dependent covariate(s) were recorded.

In this section, we use the `pbc` data (Therneau and Grambsch 2001), provided by the **survival** package (see `?pbc` and **R**-chunk 3), ignoring the potentially dependent competing risks, focusing only on the endpoint death (see also `vignette("timedep", package="survival")`). Note that by loading `pbc`, two data sets are loaded, the first, `pbc`, contains survival information and time-constant covariates (and values of time-dependent covariates recorded at beginning of the follow-up) and `pbcseq`, which stores information on time-dependent covariates.

The variables defining the structure of the data are

- the subject indicator (`id`),
- the time to event (`time`),
- the event indicator (`status`),
- the time of exposure/time at which TDCs were observed (`day`).

Note that only the first 312 observations in `pbc` also have time-dependent information in `pbcseq`, therefore we only use this part of the data.

R-chunk 3

```
# Note that this code loads two data sets, pbc and pbcseq
data("pbc", package="survival")
# event time information
pbc <- pbc %>%
  filter(id <= 312) %>%
  mutate(status = 1L*(status == 2)) %>%
  select(id:status, trt:sex, bili, protime)
pbc %>% slice(1:6)

# A tibble: 6 x 8
     id  time status   trt   age sex     bili protime
  <int> <int>  <int> <int> <dbl> <fct>  <dbl>   <dbl>
1     1   400      1     1  58.8 f     14.5      12.2
2     2  4500      0     1  56.4 f      1.10     10.6
3     3  1012      1     1  70.1 m      1.40     12.0
4     4  1925      1     1  54.7 f      1.80     10.3
5     5  1504      0     2  38.1 f      3.40     10.9
6     6  2503      1     2  66.3 f      0.800    11.0

# TDC information
pbcseq <- pbcseq %>% select(id, day, bili, protime)
pbcseq %>% slice(1:6)

# A tibble: 6 x 4
     id   day   bili protime
  <int> <int>  <dbl>   <dbl>
1     1     0 14.5      12.2
2     1   192 21.3      11.2
3     2     0  1.10     10.6
4     2   182  0.800    11.0
5     2   365  1.00     11.6
6     2   768  1.90     10.6
```

To combine these data sets and to transform them into the PED format we again use the `as_ped` function, however, the first argument is a list of data sets and the variables that should be treated as concurrent variables are specified using the `concurrent` formula special, as illustrated in **R**-chunk 4.

R-chunk 4

```
pbc_ped <- as_ped(
  data    = list(pbc, pbcseq),
  formula = Surv(time, status)~sex|concurrent(bili, protime, tz_var = "day"),
  id      = "id")
pbc_ped

# A tibble: 201,398 x 9
# Groups:   id [312]
      id tstart  tend interval  offset ped_status sex    bili protime
 * <int>  <dbl> <int> <fct>      <dbl>      <dbl> <fct> <dbl>   <dbl>
 1     1     0.    41 (0,41]     3.71          0. f      14.5    12.2
 2     1    41.    51 (41,51]    2.30          0. f      14.5    12.2
 3     1    51.    71 (51,71]    3.00          0. f      14.5    12.2
 4     1    71.    77 (71,77]    1.79          0. f      14.5    12.2
 5     1    77.   108 (77,108]   3.43          0. f      14.5    12.2
 6     1   108.   110 (108,110]  0.693         0. f      14.5    12.2
 7     1   110.   113 (110,113]  1.10          0. f      14.5    12.2
 8     1   113.   130 (113,130]  2.83          0. f      14.5    12.2
 9     1   130.   131 (130,131]  0.            0. f      14.5    12.2
10     1   131.   140 (131,140]  2.20          0. f      14.5    12.2
# ... with 201,388 more rows
```

In **R**-chunk 4 `as_ped`

- uses the union of unique event times and all measurement times of the TDCs as interval split points,
- merges the expanded data set with the data set containing information on TDCs by ID and time (`time` and `day`) and
- fills in the values of TDCs for any time-points that did not occur in `tz_var` by carrying the respective previous value of the TDC forward.

The last point of course implies the assumption that the values of the TDCs remain constant between observation points, which can be questionable, especially for longer periods between updates.

For analysis of this data and a comparison to results from an extended Cox model see Bender *et al.* (2018a) and the **pammtools** vignette on time-dependent covariates.

3.3. Time-dependent covariates with cumulative effects

Some additional effort is required to create PED with TDCs that will be modeled as cumulative effects. If `mgcv::gam` is used for estimation, we need to construct *covariate matrices* for each TDC with a cumulative effect, as well as additional covariate matrices representing either time and/or time of exposure and/or the latency of exposure and the lag-lead matrix defining the time window $\mathcal{T}(t)$.

Let's consider a model with one cumulative effect $g(\mathbf{z}, t)$ of TDC $\mathbf{z}$, such that a general representation of the cumulative effect is given by

$$g(\mathbf{z}, t) = \int_{\mathcal{T}(t)} h(t, t_z, z(t_z))\mathrm{d}t_z \tag{5}$$

In (5)

- the tri-variate function $h(t, t_z, z(t_z))$ defines the so-called *partial effects* of the TDC $z(t_z)$ observed at exposure time t_z on the hazard at time t (Bender *et al.* 2018b). Other specifications commonly used in the literature are special cases of the general partial effect definition given above, e.g.,
  - $h(t - t_z)z(t_z)$ is the WCE model of Sylvestre and Abrahamowicz (2009) and
  - $h(t - t_z, z(t_z))$ corresponds to the DLNM model of Gasparrini (2014)
- the cumulative effect $g(\mathbf{z}, t)$ at follow-up time t is the integral of the partial effects over exposure times t_z contained within $\mathcal{T}(t)$
- $\mathcal{T}(t)$ denotes the *lag-lead window* (or window of effectiveness). The most common definition is $\mathcal{T}(t) = \{t_{z,q} : t \geq t_{z,q}, q = 1, \ldots, Q\}$, which means that all exposures that were observed prior to t or at t can affect the hazard at time t.

Thus, when transforming the data to a format suitable to fit such effects using `mgcv::gam`, the required covariate matrices will be created depending on

- the specific definition of the partial effect $h()$,
- the grid of exposure times t_z and
- the lag-lead window $\mathcal{T}(t)$

As before, the `as_ped` function can be used to transform the data into the right format by extending the RHS of the `formula` using the formula special `cumulative`. The most important arguments to `cumulative` are:

`...`: a place holder where the individual components (variables) of the partial effects can be specified. See Table 2 for a selection of possible partial effect specifications and how to represent them in `cumulative` (for their specification in `mgcv::gam` see Section 4.3)

`tz_var`: the name of the variable that contains exposure times t_z of TDC $\mathbf{z}$

`ll_fun`: a boolean function of follow-up time t and exposure time t_z, which defines $\mathcal{T}(t)$ in Equation (5) (see also Figure 2)

For illustration of the data transformation using `as_ped` and `cumulative`, consider the simulated data `simdf_elra` contained in **pammtools** (see example in `?sim_pexp` for data generation):

```
data("simdf_elra", package = "pammtools")
simdf_elra %>% slice(1:3)

# A tibble: 3 x 9
     id   time status     x1    x2 tz1        z.tz1      tz2        z.tz2
  <int>  <dbl>  <int>  <dbl> <dbl> <list>     <list>     <list>     <list>
1     1  3.22       1  1.59  4.61  <int [10]> <dbl [10]> <int [11]> <dbl ~
2     2 10.0        0 -0.530 0.178 <int [10]> <dbl [10]> <int [11]> <dbl ~
3     3  0.808      1 -2.43  3.25  <int [10]> <dbl [10]> <int [11]> <dbl ~
```

It contains

- the follow-up time t (`time`),
- the event indicator (`status`, censoring only occurs at the end of the follow up at $t = 10$),
- two time constant covariates x_1 (`x1`) and x_2 (`x2`) and
- two TDCs $\mathbf{z}_1$ (`z1.tz1`), $\mathbf{z}_2$ (`z2.tz2`) observed at two different exposure time grids t_{z_1} (`tz1`) and t_{z_2} (`tz2`).

Let's further assume that two different lag-lead windows $\mathcal{T}_1(t) = \{t_{z_1,q_1} : t \geq t_{z_1,q_1}, q_1 = 1, \ldots, Q_1\}$ and $\mathcal{T}_2(t) = \{t_{z_2,q_2} : t \geq t_{z_2,q_2} + 2, q_2 = 1, \ldots, Q_2\}$ (the latter defined by `ll_2 <- function(t, tz)  t >= tz + 2`) are associated with the cumulative effects of the respective TDCs. The latter corresponds to a lag time of 2 days, so, for example, the value of $z_2(3)$ only affects the hazard for follow-up times $t \geq 5$.

Table 2 shows a selection of partial effect specifications for this setting and the respective specification using the formula special `cumulative`. Note that

- the variable representing follow-up time t in `cumulative` (here `time`) must match the time variable specified on the LHS of the formula (`Surv(time, status)`) provided to `as_ped`
- if the latency $t - t_z$ should be used instead of t_z, the variables representing exposure time t_z (here `tz1` and `tz2`) must be wrapped within `latency()`
- by default, $\mathcal{T}(t)$ is defined as `function(t, tz) t >= tz`, thus for $\mathcal{T}_1(t)$ there is no need to specify the lag-lead window explicitly. To define a custom lag-lead window, provide the respective function to the `ll_fun` argument in `cumulative` (see `ll_2` in Table 2)
- `cumulative` does not distinguish between partial effects $h(t-t_z, z(t_z))$ and $h(t-t_z)z(t_z)$ as the required data transformations are identical
- more than one `z` variable can be provided to `cumulative`, which can be convenient if multiple covariates share time components and will be integrated over the same lag-lead windows
- multiple `cumulative` terms with different exposure times t_{z_1}, t_{z_2} and/or different lag-lead windows for different covariates z_1, z_2 can be specified, as illustrated in Table 2
- to tell `cumulative` which of the variables provided is the exposure time t_z, the `tz_var` argument must be specified within each `cumulative` term. The follow-up time component t (`time`) will be recognized from the LHS of the formula

Table 2: A selection of possible partial effect specifications and the usage of `cumulative` to create matrices needed to estimate different types of cumulative effects of $\mathbf{z}_1$ and $\mathbf{z}_2$.

cumulative effect(s)	data transformation (`pammtools`)
$\int_{\mathcal{T}_1} h(t - t_{z_1}, z_1(t_{z_1}))$	`cumulative(latency(tz1), z1.tz1, tz_var="tz1")`
$\int_{\mathcal{T}_1} h(t, t - t_{z_1}, z_1(t_{z_1}))$	`cumulative(time, latency(tz1), z1.tz1, tz_var="tz1")`
$\int_{\mathcal{T}_1} h(t, t_{z_1}, z_1(t_{z_1}))$	`cumulative(time, tz1, z1.tz1, tz_var="tz1")`
$\int_{\mathcal{T}_1} h(t, t_{z_1}, z_1(t_{z_1})) + \int_{\mathcal{T}_2} h(t - t_{z_2}, z_2(t_{z_2}))$	`cumulative(time, tz1, z1.tz1, tz_var="tz1") + cumulative(latency(tz2), z2.tz2, tz_var="tz2", ll_fun=ll_2)`

One possible data transformation call for the `simdf_elra` data is given in **R**-chunk 5.

R-chunk 5

```
ped_simdf <- simdf_elra %>% as_ped(Surv(time, status)~ x1 + x2|
    cumulative(time, latency(tz1), z.tz1, tz_var="tz1") +
    cumulative(latency(tz2), z.tz2, tz_var="tz2"), cut = 0:10)
str(ped_simdf)
```

```
...
 $ tend        : int  1 2 3 4 1 2 3 4 5 6 ...
 $ interval    : Factor w/ 10 levels "(0,1]","(1,2]",..: 1 2 3 4 1 2 3 4 5 6 ...
 $ offset      : num  0 0 0 -1.53 0 ...
 $ ped_status  : num  0 0 0 1 0 0 0 0 0 0 ...
 $ x1          : num  1.59 1.59 1.59 1.59 -0.53 ...
 $ x2          : num  4.612 4.612 4.612 4.612 0.178 ...
 $ time_tz1_mat: int [1:1004, 1:10] 0 1 2 3 0 1 2 3 4 5 ...
 $ tz1_latency : num [1:1004, 1:10] 0 0 1 2 0 0 1 2 3 4 ...
 $ z.tz1_tz1   : num [1:1004, 1:10] -2.014 -2.014 -2.014 -2.014 -0.978 ...
 $ LL_tz1      : num [1:1004, 1:10] 0 1 1 1 0 1 1 1 1 1 ...
 $ tz2_latency : num [1:1004, 1:11] 5 6 7 8 5 6 7 8 9 10 ...
 $ z.tz2_tz2   : num [1:1004, 1:11] -0.689 -0.689 -0.689 -0.689 0.693 ...
 $ LL_tz2      : num [1:1004, 1:11] 1 1 1 1 1 1 1 1 1 1 ...
...
```

The newly created matrix valued variables have

- different number of columns (10 vs. 11), reflecting the different exposure time grids ($t_{z_1,1}, \ldots, t_{z_1,Q_1=10}$ and $t_{z_2,1} = -5, \ldots, t_{z_2,Q_2} = 5$).
- different components, depending on the partial effect and `cumulative` specification, respectively. Thus, for `z.tz1` a time matrix `time_tz1` was created as well as a latency matrix `tz1_latency`, whereas only the latency matrix `tz2_latency` was created for the partial effects associated with `z.tz2`.
- different lag-lead specifications, which can be extracted and visualized using convenience functions `get_laglead` and `gg_laglead`. Applied to a `ped` object, they retrieve the lag-lead definition used during data transformation (cf. Figure 1). More complex specifications of $\mathcal{T}(t)$ can be generated easily (cf. Figure 2), where a lead time of $t_{\text{lead}} = 5$ is included in addition to a lag time of $t_{\text{lag}} = 2$.

```
gg_laglead(ped_simdf)
```

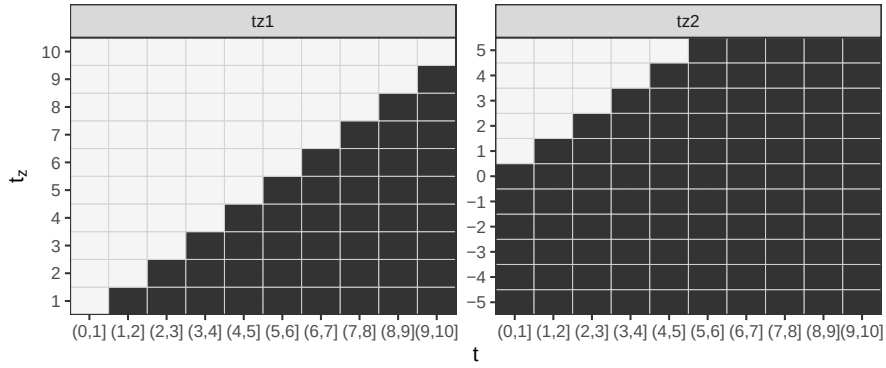


Figure 1: Lag-lead windows created by `as_ped` in **R**-chunk 5. When viewed row-wise, the black squares indicate the intervals at which the respective exposure times t_z can affect the hazard. For example, in the left panel, exposure at time $t_z = 5$ can affect the hazard in intervals $(5, 6]$ through $(9, 10]$ (`as_ped` is conservative and $t \geq t_z$ is only true if the relationship is true for the interval start time). When viewed column-wise, one can obtain the exposure times contained within $\mathcal{T}(t)$. For example, $\mathcal{T}(t = 5) = \mathcal{T}((\kappa_{j-1} = 4, \kappa_j = 5]) = \{t_z = 1, \ldots, t_z = 4\}$.

```
my_ll_fun <- function(t, tz, tlag = 2, tlead = 5) {
  t >= tz + tlag & t < tz + tlag + tlead
}
gg_laglead(0:10, tz=-5:5, ll_fun = my_ll_fun)
```

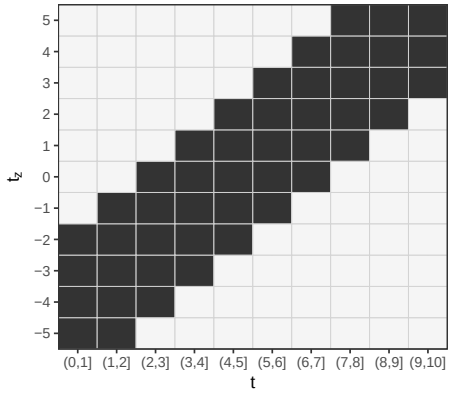


Figure 2: Illustration of a more complex definition of the lag-lead window $\mathcal{T}(t)$ with $t_{\text{lag}} = 2$ and $t_{\text{lead}} = 5$. For example, exposure at time $t_z = -1$, starts to affect the hazard at time $t = t_z + t_{\text{lag}} = -1 + 2 = 1$, i.e., interval $(1, 2]$, as t in the specification of the lag-lead function refers to the start time of the interval. Similarly, exposure at time t_z lasts until $t = t_z + t_{\text{lag}} + t_{\text{lead}} = -1 + 2 + 5 = 6$, i.e., interval $(5, 6]$. Note that we used the condition $t < tz + t_{\text{lag}} + t_{\text{lead}}$ to ensure that the condition is true for the end time of the interval.

4. Modeling and Interpretation

With data in PED format (see Section 3), the subsequent modeling step is relatively straightforward, as any software for Generalized Additive (Mixed) Models (or similar) can be used. In this article, the model estimation is performed outside the **pammtools** package using **mgcv** (Wood 2011). In the following sections, we demonstrate how to fit different models using the `mgcv::gam` formula syntax, with special attention given to cumulative effects.

4.1. Time-constant effects

We start with a standard survival model with time-constant effects of time-constant covariates and compare the results to the Cox PH model using the tumor data (`?tumor`) contained in the **pammtools** package.

The data used in this section has already been transformed into the correct format in Section 3.1 (see **R**-chunk 2). Therefore, we can directly apply `mgcv::gam` to the transformed data as shown in **R**-chunk 6. Note that we must specify `family = poisson()` and `offset = offset` for the model to be estimated correctly. For an overview of estimates the **mgcv** functions `summary.gam` and `plot.gam` can be used. Note that the log-baseline hazard displayed in Figure 3 does not contain the intercept term β_0 and cannot be interpreted usefully as it relates to a patient with age 0. Note that `gg_smooth` replicates the plots produced by `plot.gam` and visualizes all effects as smooth lines, while for PAMMs, representations of the (log-)hazard should be plotted as step functions (see Figure 4).

R-chunk 6

```
pam_tumor <- gam(
  formula = ped_status ~ s(tend) + sex + age + charlson_score + transfusion +
    + complications + metastases + resection,
  data = ped_tumor, family = poisson(), offset = offset, method = "REML")
# default summary
summary(pam_tumor)

...

Parametric coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -9.837979   0.364656 -26.979  < 2e-16 ***
sexfemale         0.185245   0.107953   1.716 0.086167 .
age               0.021019   0.005034   4.175 2.98e-05 ***
charlson_score    0.149562   0.041992   3.562 0.000368 ***
transfusionyes    0.254105   0.110703   2.295 0.021711 *
complicationsyes  0.581987   0.109125   5.333 9.65e-08 ***
metastasesyes     0.166650   0.116752   1.427 0.153467
resectionyes      0.260660   0.112118   2.325 0.020079 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
          edf Ref.df Chi.sq p-value
s(tend) 3.761  4.679  19.33 0.00139 **
...
```

```
gg_smooth(ped_tumor, pam_tumor, terms="tend") + xlab("time")
```

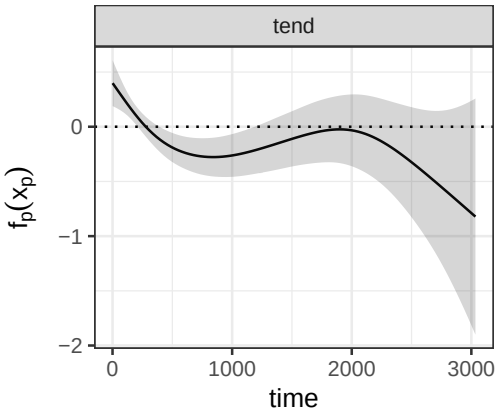


Figure 3: Log-baseline hazard of the PAM estimated on the tumor data with time-constant effects (cf. **R**-chunk 6).

pammtools provides convenience functions to extract the fixed coefficients including confidence intervals (`tidy_fixed`, cf. **R**-chunk 7) as well as a plot function for the fixed effect coefficients (`?gg_fixed`), which returns a `ggplot` object. Note that by default, the output of both functions omits the intercept term, which can be added by setting `intercept=TRUE`. When comparing the results with the Cox PH model (cf. **R**-chunk 7), the estimated effects are, not surprisingly, very similar.

R-chunk 7

```
coxph_tumor <- coxph(
  formula = Surv(days, status) ~ sex + age + charlson_score +transfusion +
    + complications + metastases + resection,
  data    = tumor)
# compare coefficient estimates
imap(list(PAM = pam_tumor, COX = coxph_tumor),
    ~ tidy_fixed(.x) %>% select(variable, coef) %>% rename(!!.y := coef)) %>%
  reduce(left_join)

# A tibble: 7 x 3
  variable            PAM    COX
  <chr>             <dbl>  <dbl>
1 sexfemale        0.185  0.185
2 age              0.0210 0.0209
3 charlson_score   0.150  0.147
4 transfusionyes   0.254  0.255
5 complicationsyes 0.582  0.571
6 metastasesyes    0.167  0.164
7 resectionyes     0.261  0.256
```

4.2. Time-varying effects

Time-varying effects of time-constant covariates $f(t) \cdot x$ can generally be divided in two groups:

- stratified hazards for categorical x
- time-varying coefficients for continuous x

Interactions between continuous and categorical covariates are possible as well in order to allow for the time-varying effect of a continuous variable to vary over the different levels of a categorical variable.

Stratified hazards model

Consider the variable `complications` for the case of stratified hazards. Suppose that patients experiencing major complications during surgery are under increased risk immediately afterwards, and that this increase subsides after some time. If this is the case, the PH assumption of the Cox model is not fulfilled, or more generally, the effect of `complications` is time-varying. One solution to this problem are stratified hazards models (e.g., Klein and Moeschberger (1997, Ch. 9.3)) with separate baseline hazards for the levels of a categorical covariate. The estimated log-hazards are presented in **R**-chunk 8 and Figure 4. Note that we use `tidy_smooth` to extract the data used by `plot.gam` for visualization of 1D smooth effects. The hazards in the two groups are vastly different with the expected drop in the log-hazard within the first 500 days for patients with major complications.

R-chunk 8

```
pam_strata <- bam(
  formula = ped_status ~  complications + s(tend, by = complications) + sex +
    age + charlson_score + transfusion + metastases + resection,
  data = ped_tumor, family = poisson(), offset = offset, discrete = TRUE)
summary(pam_strata)

...

Parametric coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -9.959335   0.363745 -27.380  < 2e-16 ***
complicationsyes  0.443763   0.122720   3.616 0.000299 ***
sexfemale         0.190760   0.108295   1.761 0.078157 .
age               0.020753   0.005018   4.136 3.53e-05 ***
charlson_score    0.159937   0.042035   3.805 0.000142 ***
transfusionyes    0.234964   0.111398   2.109 0.034924 *
metastasesyes     0.175349   0.116637   1.503 0.132744

Approximate significance of smooth terms:
                          edf Ref.df Chi.sq p-value
s(tend):complicationsno  4.434  5.481  11.05  0.0746 .
s(tend):complicationsyes 5.087  6.181  91.59  <2e-16 ***
...
```

```
tidy_smooth(pam_strata) %>%
  ggplot(aes(x = x, y = fit)) +
    geom_stepribbon(aes(ymin = ci_lower, ymax = ci_upper), alpha = 0.3) +
    geom_step() + geom_hline(yintercept = 0, lty = 2) +
    facet_wrap(~ylab) +
    xlab(expression(t)) + ylab(expression(f[p](t) %.% x[p]))
```

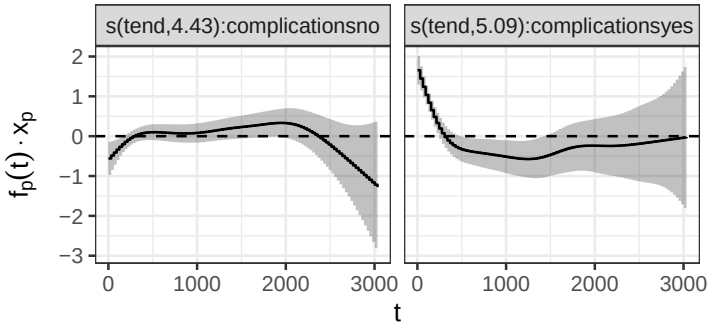


Figure 4: Stratified log-hazards for patients with (right) and without (left) major complications.

Varying coefficients

Let's now include all covariates available in the `tumor` data, with possibly non-linearly time-varying effects, where the effects of continuous covariates are assumed to vary non-linearly in time, but linearly in the covariate, i.e., $f_p(t)x_p$. The model specification is given in **R**-chunk 9. Note that categorical covariates are included using `by = as.ordered(...)`, which (together with `ti`) ensures identifiability of the model (cf. `?mgcv::gam.models` and `?mgcv::ti`). For the effects of `age` and `charlson_score` the basis functions of the smooths are multiplied with the respective covariate values, thus no further identifiability constraints are necessary.

R-chunk 9

```
pam_tumor_tve <- bam(
  formula = ped_status ~ ti(tend) +
    complications + ti(tend, by = as.ordered(complications)) +
    metastases    + ti(tend, by = as.ordered(metastases))    +
    sex           + ti(tend, by = as.ordered(sex))           +
    transfusion   + ti(tend, by = as.ordered(transfusion))   +
    resection     + ti(tend, by = as.ordered(resection))     +
    s(tend, by = charlson_score) +
    s(tend, by = age),
  data   = ped_tumor, family = poisson(), offset = offset,
  method = "fREML", discrete = TRUE)
```

The model output is presented in **R**-chunk 10. The effects of variables `metastases`, `transfusion` and `resection` were estimated as linearly time-varying effects (edf=1), however, they must be interpreted as relative changes (*ceteris paribus*, c.p.) compared to the baseline hazard `ti(tend)`, which itself is non-linear.

R-chunk 10

```
summary(pam_tumor_tve)

...
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -9.9354     0.3623 -27.425  < 2e-16 ***
complicationsyes   0.3775     0.1230   3.070  0.00214 **
metastasesyes      0.2156     0.1183   1.822  0.06847 .
sexfemale          0.2138     0.1084   1.973  0.04855 *
transfusionyes     0.2037     0.1154   1.765  0.07757 .
resectionyes       0.2820     0.1134   2.487  0.01287 *
---
Approximate significance of smooth terms:
                                       edf Ref.df Chi.sq  p-value
ti(tend)                             1.360  1.608  2.093 0.180850
ti(tend):as.ordered(complications)yes 3.703  3.931 88.476  < 2e-16 ***
ti(tend):as.ordered(metastases)yes    1.000  1.001 12.012 0.000531 ***
ti(tend):as.ordered(sex)female        1.866  2.259  3.859 0.185534
ti(tend):as.ordered(transfusion)yes   1.000  1.000  2.355 0.124894
ti(tend):as.ordered(resection)yes     1.000  1.000  2.793 0.094669 .
s(tend):charlson_score                2.000  2.000 23.017 1.01e-05 ***
s(tend):age                           2.000  2.000 14.656 0.000657 ***
...
```

The usual visualization of the log-hazard contributions $f_p(t)x_p$ over the follow-up could be used for the interpretation of the estimates (similar to figure 4). However, for models with time-varying effects (that are linear in the covariates), an alternative visualization, which is also useful for comparisons to the non-parametric additive Aaalen model (Martinussen and Scheike 2006), will be used here.

The default visualization of covariate effect estimates for the Aalen model in the **timereg** package is the so-called cumulative coefficient $\mathcal{B}_p(t) = \int_0^t \beta_p(s)\mathrm{d}s$. Since the Aalen model is additive, i.e., $\lambda(t|\mathbf{x}) = \lambda_0(t) + \beta_1(t)x_1(t) + \cdots$, this cumulative coefficient can be nicely interpreted as the cumulative hazard difference at time t for a 1 unit increase of the covariate/compared to its reference level (c.p.), i.e., $\mathcal{B}(t) = \Lambda(t|x+1) - \Lambda(t|x)$. Thus, to obtain a PAMM analog of the cumulative coefficient, we can calculate the difference between the respective cumulative hazards. Although $\mathcal{B}(t)$ is not directly estimated for PAMMs as it is for the Aalen model, **pammtools** provides the function `get_cumu_coef` that performs these calculations (including simulation based confidence intervals), as illustrated in **R**-chunk 11.

The cumulative coefficients of the PAMM and Aalen model are presented in Figure 5. The cumulative hazard difference between a patient with complications (compared to one without, c.p.), increases at the beginning, directly after the operation when complications occurred, while after approximately 500 days, the cumulative hazard difference remains constant (i.e. $\beta_p(t) = f_p(t) \approx 0 \,\forall\, t > 500$). Similarly, the effect of metastases has a plausible interpretation: At $t = 0$, as much as possible of the cancerous tissue including metastases is removed, thus the hazard in both groups is almost the same in the beginning, however, the risk of cancer

returning after some time due to cancerous tissue that was not removed is higher in patients with metastases, which notably increases their hazard for $t > 1500$ compared to patients without metastases. For the cumulative coefficients based on PAMMs, confidence intervals were estimated by Monte Carlo estimation based on 100 draws from the model coefficients' posterior distribution (Argyropoulos and Unruh 2015; Wood 2017). Overall, the estimates obtained from the PAMM estimates are very close to the estimates obtained from the Aalen model with respect to the cumulative coefficients as well as their confidence intervals.

R-chunk 11

```
# here cumu_hazard denotes the cumulative hazard differences
get_cumu_coef(pam_tumor_tve, ped_tumor, terms = c("age", "sex")) %>%
  group_by(variable) %>% slice(1:2)

# A tibble: 4 x 6
# Groups:   variable [2]
  method variable      time cumu_hazard   cumu_lower cumu_upper
  <chr>  <chr>        <dbl>       <dbl>        <dbl>      <dbl>
1 bam    age             1.  0.00000458  0.000000896 0.00000883
2 bam    age             2.  0.00000916  0.00000180  0.0000177
3 bam    sex (female)    1. -0.0000177  -0.000116    0.0000858
4 bam    sex (female)    2. -0.0000352  -0.000231    0.000172
```

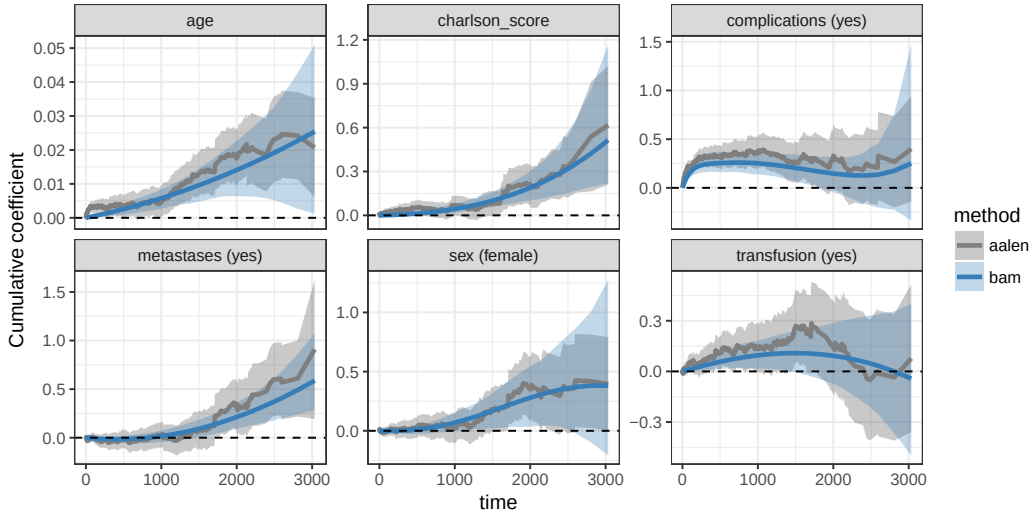


Figure 5: Comparison of cumulative coefficients estimated with PAMMs and the additive Aalen model respectively (the effect of resection is not displayed for conciseness). For PAMMs these are defined as cumulative hazard differences, e.g. $\mathcal{B}_{\text{PAMM}}(t) := \Lambda(t|\text{sex} = \text{"female"}) - \Lambda(t|\text{sex} = \text{"male"})$. Note the different scales on the vertical axes of the panels.

4.3. Cumulative effects

In this section, we illustrate the estimation of cumulative effects using `mgcv::gam` (or `mgcv::bam`) with suitably formatted data sets (see Section 3.3), as well as their visualization. We use simulated data that allows us to discuss different aspects and model classes covered by our general approach. The simulation of the various data sets with different specifications of cumulative effects is described in Appendix A, specifically sections A.3.1, A.3.2 and A.3.3

Weighted cumulative exposure

Consider model (6) with a smooth log-baseline hazard function $f_0(t)$ and a cumulative covariate effect of exposure histories $\mathbf{z}_i$. In the following example, the associated partial effect is non-linear in the latency $t - t_z$, the time since the exposure was observed, and linear in the values of $z(t_z)$, such that

$$\lambda_i(t|\mathbf{z}_i) = \exp\left(\beta_0 + f_0(t) + 0.5x_{1,i} + \sqrt{x_{2,i}} + \int_{\mathcal{T}(t)} h(t - t_z)z_i(t_z)\mathrm{d}t_z\right) \tag{6}$$

Section A.3.1 describes how to simulate data from this model using the **pammtools** function `sim_pexp` (cf. **R**-chunk 19). Given this data (`simdf_wce`), we can proceed with the analysis of the data, first by transforming it to the PED format using the `as_ped` function as shown in Section 3.3 and applied to the simulated data in **R**-chunk 12. Note that the created matrix columns have 41 columns, because this was the length of the exposure time grid used in the data simulation step.

R-chunk 12

```
time_grid <- seq(0, 10, by = 0.5)
ped_wce <- as_ped(
  data = simdf_wce,
  formula = Surv(time, status) ~ x1 + x2|
    cumulative(latency(tz), z.tz, tz_var="tz", ll_fun = ll_fun),
  cut = time_grid)
str(ped_wce,1)
```

```
...
 $ tz_latency: num [1:7460, 1:41] 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 ...
 $ z.tz      : num [1:7460, 1:41] 1.86 1.86 1.86 1.86 1.86 ...
 $ LL        : num [1:7460, 1:41] 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 ...
...
```

R-chunk 13 shows the model specification necessary to fit the correctly specified model. Note that we use the correct lag-lead window, as we provide the true `ll_fun` (cf. **R**-chunk 18) to the data transformation function in **R**-chunk 12. The estimated weight function $\hat{h}(t - t_z)$ is fairly close to the true function used in the simulation, as displayed in Figure 6.

R-chunk 13

```
mod_wce <- gam(
  formula = ped_status ~ s(tend) + s(x1) + s(x2) + s(tz_latency, by = z.tz * LL),
  data    = ped_wce, family  = poisson(), offset  = offset, method  = "REML")
summary(mod_wce)

...

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.77996    0.04739  -37.56   <2e-16 ***

Approximate significance of smooth terms:
                         edf Ref.df Chi.sq  p-value
s(tend)                6.366  7.385 328.49  < 2e-16 ***
s(x1)                  1.420  1.728 449.38  < 2e-16 ***
s(x2)                  3.021  3.758 199.97  < 2e-16 ***
s(tz_latency):z.tz * LL 3.566  4.182  43.46 1.23e-08 ***
...
```

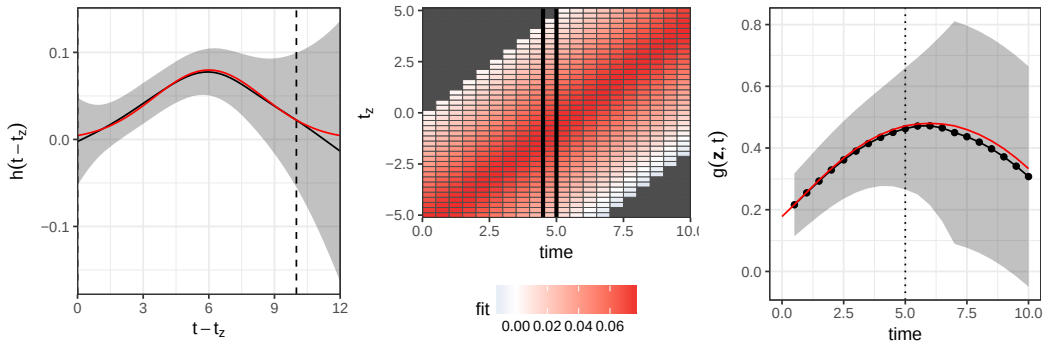


Figure 6: Left: Partial effect $\hat{h}(t - t_z)$ estimated in **R**-chunk 13, depicted for all possible latencies for the particular data. Dashed lines indicate the latencies that contribute to the cumulative effect at interval $(4.5, 5]$. Middle: Partial effects for each combination of t and t_z. The vertical stripes at each interval are subsets of the partial effect depicted in the left panel. Right: Cumulative effect $g(\mathbf{z}, t)$ at all time points of the follow up. Each point is the sum of the vertical stripes depicted in the middle panel. The point at $t = 5$ indicates the sum of weighted partial effects of the highlighted vertical stripe (interval $(4.5, 5]$) in the middle panel.

Distributed Lag Non-linear Model

The WCE approach from the previous section assumes that the effect of $\mathbf{z}$ is non-linear with respect to the latency and linear in $\mathbf{z}$. Relaxing the latter assumption and allowing the partial effect to also vary non-linearly over $z(t_z)$ (cf. eq. (9)) leads to what is often referred to as

the distributed lag non-linear model (DLNM; Gasparrini 2014).

$$\lambda_i(t|\mathbf{z}_i) = \exp\left(\beta_0 + f_0(t) + 0.5x_{1,i} + \sqrt{x_{2,i}} + \int_{\mathcal{T}(t)} h(t - t_z, z_i(t_z))\mathrm{d}t_z\right) \tag{7}$$

Data transformation and model estimation for this data (`simdf_dlnm`; cf. Section A.3.2 for data simulation and Figure 14 for the true partial effects used for simulation) is given in **R**-chunk 14. Note that the formula provided to `as_ped` is actually the same as the one used to transform the `simdf_wce` data in **R**-chunk 12, as the created covariate matrix for `z.tz` will be the same in both cases, thus we could have also used the `ped_wce` data for estimation of the DLNM model. However, the specification of the term in the call to `gam` is different: `te(tz_latency, z.tz, by = LL)` for the DLNM vs. `s(tz_latency, by = z.tz * LL)` for the WCE.

R-chunk 14

```
ped_dlnm <- as_ped(
  formula = Surv(time, status) ~ x1 + x2|
    cumulative(latency(tz), z.tz, tz_var = "tz", ll_fun = ll_fun),
  data = simdf_dlnm, cut = time_grid)
# ped_dlnm$tz_latency <- ped_dlnm$tz_latency * ped_dlnm$LL
mod_dlnm <- bam(
  formula  = ped_status ~ s(tend) + s(x1) + s(x2) +
    te(tz_latency, z.tz, by = LL, k = c(10,10)),
  data = ped_dlnm, family = poisson(), offset = offset,
  method = "fREML", discrete = TRUE)
summary(mod_dlnm)


...
te(tz_latency,z.tz):LL 8.795 11.424  46.26 4.42e-06 ***
---
...
```

Figure 7 depicts the estimated partial effect surface (left hand panel) as well as one-dimensional slices through the surface with respect to the latency $t - t_z \in \{1, 5, 10\}$ (middle panel) and the covariate $z(t_z) \in \{-1.5, 0, 1.5\}$ (right panel). Note that, equivalently to the true partial effect in Figure 14, the depicted effects are relative to an observation with exposure history $z(t_z) = -1\,\forall\, t_z$, thus the effects pass through zero at $z(t_z) = -1\,\forall\, t, t_z$. We use **pammtools** convenience functions `gg_partial` and `gg_slice` to create the individual figures. Internally, they use `make_newdata` to create a data set based on `ped_dlnm` and the variable specification provided through the ellipsis arguments (`...`). If specified, the effects will be calculated relative to covariate values provided as the `reference` argument (here `reference = list(z.tz = -1)`), which must be a list with single value specifications for each covariate that should be changed in the comparison data set.

Figure 8 again shows the partial effect surface from Figure 7 (left panel), as well as the partial effects for each combination of t and t_z, with $z(t_z) = 1 \forall t_z$. This visualization shows more directly which partial effects will contribute to the cumulative effect at time t (see also the dashed lines in the left panel). Finally, the right panel of Figure 8 depicts the total cumulative effect $g(\mathbf{z}, t)$ for the partial effects displayed in the middle panel.

```
# define reference values
ref  <- list(z.tz = -1)
# partial effect surface
p_partial_dlnm <- gg_partial(ped_dlnm, mod_dlnm, term = "z.tz", reference = ref,
  z.tz = seq(-3, 3, by = 0.1), tz_latency = seq(0, 12, by = .25), LL=c(1))
# slices over exposures with fixed exposure time values
p_slice_tz <- gg_slice(ped_dlnm, mod_dlnm, term = "z.tz", reference = ref,
  z.tz = seq(-3, 3, by = 0.25), tz_latency = c(1, 5, 10), LL = c(1)) +
  geom_vline(xintercept = 1.5, lty = 3)
# slices over exposure times with fixed exposure values
p_slice_z.tz <- gg_slice(ped_dlnm, mod_dlnm, term = "z.tz", reference = ref,
  z.tz = c(-1.5, 0, 1.5), tz_latency = seq(0, 12, by = 0.25), LL = c(1)) +
  geom_vline(xintercept = 6, lty = 3) +
  scale_colour_brewer(palette = "Dark2")+ scale_fill_brewer(palette = "Dark2")
```

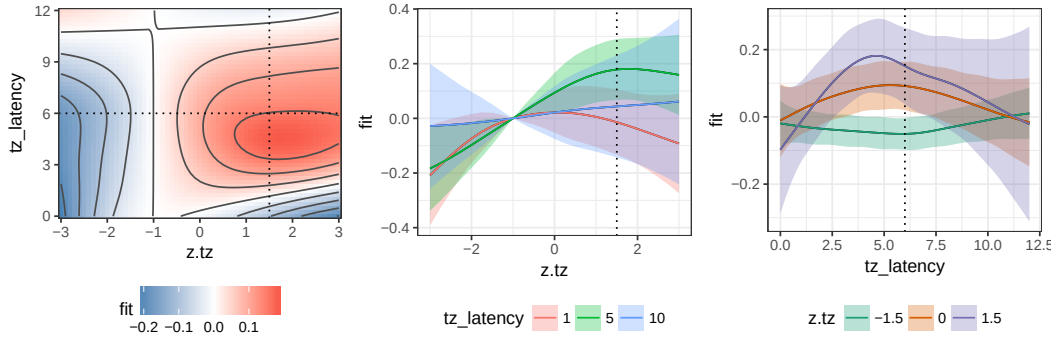


Figure 7: Partial effect $\hat{h}(t - t_z, z(t_z))$ estimated by model `mod_dlnm` in **R**-chunk 14. Note, all effects were calculated relative to $z(t_z) = -1 \forall t_z$. Left: Partial effect surface for a range of values for latency $t - t_z$ and covariate $z(t_z)$. Middle: Slices through partial effect surface for latencies 1, 5 and 10. Right: Slices through the partial effect surface for $z(t_z) \in \{-1.5, 0, 1.5\}$.

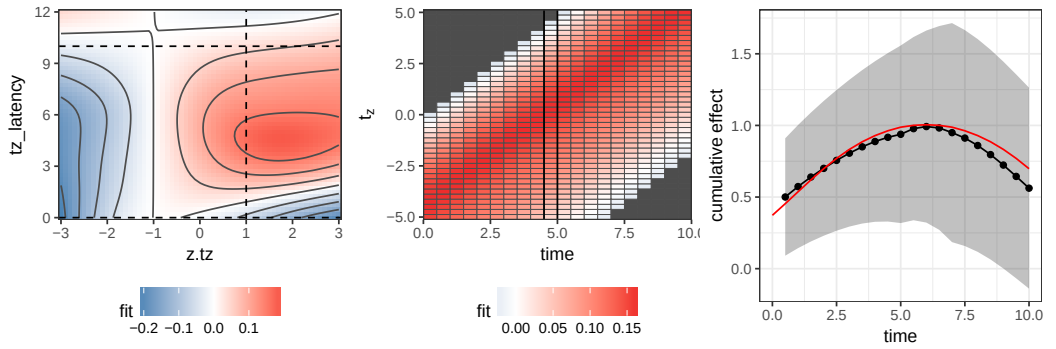


Figure 8: From left to right: Bivariate partial effect surface estimate $h(t - t_z, z(t_z))$, partial effects for different combinations of t and t_z with $z(t_z) = 1, \forall t_z$ and the resulting cumulative effect $g(\mathbf{z}, t)$.

General Exposure-lag-response Associations

In Sections 4.3.1 and 4.3.2 we discussed the most common specifications of cumulative effects in the literature. Our general specification of cumulative effects in eq. (5) has the advantage that it includes the other approaches as special cases and while also supporting alternative (and more complex) models. Thus, depending on the context, alternative specifications of the partial effects are possible, e.g.,

- $h(t, t - t_z)z(t_z)$ or alternatively $h(t, t_z)z(t_z)$, a smoothly time-varying WCE (the latter formulation was used in Bender *et al.* (2018b) in combination with a categorical $z(t_z)$)
- $h(t, t - t_z, z(t_z))$, a smoothly time-varying DLNM, which was demonstrated by means of a simulation study in Bender *et al.* (2018b, sec. 4)

For a last illustration, consider the following model:

$$\lambda_i(t|\mathbf{z}_i) = \exp\left(\beta_0 + f_0(t) + 0.5x_{1,i} + \sqrt{x_{2,i}} + \int_{\mathcal{T}(t)} h(t, t_z)z_i(t_z)\mathrm{d}t_z\right) \tag{8}$$

which looks very similar to the WCE model in Section 4.3.1, but the assumption that the partial effect only depends on the latency $t - t_z$ is softened. Data simulation from model (8) is given in **R**-chunk 21 and the true bivariate partial effect $h(t, t_z)$ as well as the resulting cumulative effect $\int_{\mathcal{T}(t)} h(t, t_z)z(t_z)\mathrm{d}t_z$ are depicted in Figure 15.

The data transformation and model estimation for this data is shown in **R**-chunk 15. The estimated effects are visualized in Figure 9. Although the bivariate partial effect surface (left panel) was estimated quite well, there is some underestimation for $t > 5$, thus, necessarily, the cumulative effect (right panel) for $t > 5$ is also underestimated.

R-chunk 15

```
# transform simulated data to PED format
ped_tv_wce <- as_ped(Surv(time, status)~ x1 + x2|
    cumulative(time, tz, z.tz, tz_var = "tz", ll_fun = ll_fun),
  data = simdf_tv_wce, cut = time_grid)

# estimate the model
mod_tv_wce <- gam(ped_status ~ s(tend) + s(x1) + s(x2) + te(time_mat, tz, by = z.tz*LL),
  data = ped_tv_wce, family = poisson(), offset = offset, method = "REML")
summary(mod_tv_wce)


...
Approximate significance of smooth terms:
                            edf Ref.df Chi.sq p-value
s(tend)                   6.726  7.754  267.0  <2e-16 ***
s(x1)                     1.002  1.004  320.7  <2e-16 ***
s(x2)                     2.689  3.351  169.5  <2e-16 ***
te(time_mat,tz):z.tz * LL 10.856 13.482  176.8  <2e-16 ***
...
```

```
# partial effect (in lag-lead window)
p_partial_elra <- gg_partial_ll(ped_tv_wce, mod_tv_wce, term="z.tz",
  time_mat = seq(0,10, by = 0.5), tz = seq(-5, 5, by = 0.25), z.tz=c(1),
    reference = list(time_mat = c(5)), time_var = "time_mat")+
  geom_contour(aes(z = fit), color = "grey30")
# cumulative effect
p_cumu_elra <- gg_cumu_eff(ped_tv_wce, mod_tv_wce, term = "z.tz", z1=1) +
  geom_line(data=cumu_df_elra, aes(x =t, y = cumu_eff), col = 2)
gridExtra::grid.arrange(p_partial_elra, p_cumu_elra, nrow=1, widths=c(1.5, 1))
```

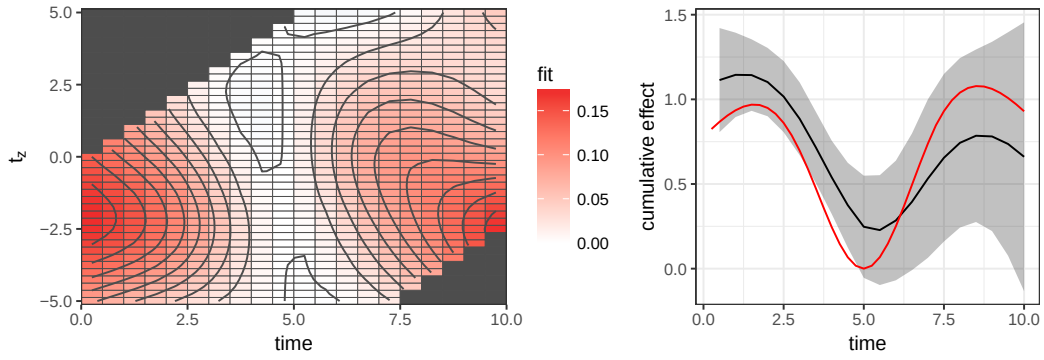


Figure 9: Left: Estimated bivariate partial effect surface $\hat{h}(t, t_z)$ for all combinations of t and t_z within $\mathcal{T}(t)$. Right: Resulting cumulative effect estimation for $z(t_z) = 1 \forall t_z$.

4.4. Convenience functions, survival probabilities and other quantities

For communicating and checking the results of complex time-to-event models, it is often necessary to calculate covariate effects in terms of conditional hazards, cumulative hazards or survival probabilities. **pammtools** provides convenience functions to quickly calculate these quantities for different covariate specifications, along with uncertainty estimates. The suggested workflow for these calculations is to create a dataset with the covariate specifications of interest and then use one of the `add_*` functions (see `?add_hazard` for an overview). For illustration we will use the `tumor` data model discussed in section 4.2.

Creating new data

pammtools provides several functions that facilitate the creation of data sets with customized covariate specifications:

- `int_info` provides interval information (start and stop times, interval length) for a given interval split point specification or extracting the split-points used during the creation of a `ped` object

```
# extract interval information
int_info(ped_tumor) %>% slice(1:5)


# A tibble: 5 x 5
  tstart  tend intlen intmid interval
```

```
   <dbl> <dbl>  <dbl>  <dbl> <fct>
1     0.    1.     1.  0.500 (0,1]
2     1.    2.     1.  1.50  (1,2]
3     2.    3.     1.  2.50  (2,3]
4     3.    5.     2.  4.00  (3,5]
5     5.    6.     1.  5.50  (5,6]
```

- `sample_info` extracts the mean and modal values for continuous and categorical variables respectively (if applied to an object of class `ped`, variables representing interval information are omitted)

```
# sample means/modi
sample_info(tumor)


# A tibble: 1 x 9
   days status charlson_score   age sex   transfusion complications
  <dbl>  <dbl>          <dbl> <dbl> <fct> <fct>       <fct>
1 1017.  0.483           2.78  62.0 male  no          no
# ... with 2 more variables: metastases <fct>, resection <fct>


sample_info(ped_tumor)


# A tibble: 1 x 7
  charlson_score   age sex   transfusion complications metastases
           <dbl> <dbl> <fct> <fct>       <fct>         <fct>
1           2.78  62.0 male  no          no            yes
# ... with 1 more variable: resection <fct>


ped_tumor %>% group_by(sex) %>% sample_info()


# A tibble: 2 x 7
# Groups:   sex [2]
  charlson_score   age sex    transfusion complications metastases
           <dbl> <dbl> <fct>  <fct>       <fct>         <fct>
1           2.96  63.3 male   no          no            yes
2           2.52  60.1 female no          no            yes
# ... with 1 more variable: resection <fct>
```

- `ped_info` combines `int_info` and `sample` info to return a data frame with all unique intervals of the `ped` object and all covariates set to their sample mean/modus.

```
# interval and sample info
ped_info(ped_tumor) %>% slice(1:3)


# A tibble: 3 x 12
  tstart  tend intlen intmid interval charlson_score   age sex
   <dbl> <dbl>  <dbl>  <dbl> <fct>             <dbl> <dbl> <fct>
1     0.    1.     1.  0.500 (0,1]              2.78  62.0 male
2     1.    2.     1.  1.50  (1,2]              2.78  62.0 male
```

```
3     2.    3.     1.  2.50  (2,3]              2.78  62.0 male
# ... with 4 more variables: transfusion <fct>, complications <fct>,
#   metastases <fct>, resection <fct>

ped_tumor %>% group_by(sex) %>% ped_info() %>% slice(1:3)

# A tibble: 6 x 12
# Groups:   sex [2]
  tstart  tend intlen intmid interval charlson_score   age sex
   <dbl> <dbl>  <dbl>  <dbl> <fct>             <dbl> <dbl> <fct>
1     0.    1.     1.  0.500 (0,1]              2.96  63.3 male
2     1.    2.     1.  1.50  (1,2]              2.96  63.3 male
3     2.    3.     1.  2.50  (2,3]              2.96  63.3 male
4     0.    1.     1.  0.500 (0,1]              2.52  60.1 female
5     1.    2.     1.  1.50  (1,2]              2.52  60.1 female
6     2.    3.     1.  2.50  (2,3]              2.52  60.1 female
# ... with 4 more variables: transfusion <fct>, complications <fct>,
#   metastases <fct>, resection <fct>
```

- `make_newdata` is a flexible function for creating new data sets from `ped` or `data.frame`-objects. Specific covariate values can be provided through the ellipsis argument (`...`) as key-value-pairs, while all unspecified variables will be set to their sample means or modes.

```
# make arbitrary new data
make_newdata(tumor, age=seq_range(age, n=3))

# A tibble: 3 x 9
   days status charlson_score   age sex   transfusion complications
  <dbl>  <dbl>          <dbl> <dbl> <fct> <fct>       <fct>
1 1017.  0.483           2.78   14. male  no          no
2 1017.  0.483           2.78   55. male  no          no
3 1017.  0.483           2.78   96. male  no          no
# ... with 2 more variables: metastases <fct>, resection <fct>

tumor %>%
  make_newdata(age=seq_range(age, n=3), sex = unique(sex), resection=c("yes"))

# A tibble: 6 x 9
   days status charlson_score   age sex    transfusion complications
  <dbl>  <dbl>          <dbl> <dbl> <fct>  <fct>       <fct>
1 1017.  0.483           2.78   14. female no          no
2 1017.  0.483           2.78   55. female no          no
3 1017.  0.483           2.78   96. female no          no
4 1017.  0.483           2.78   14. male   no          no
5 1017.  0.483           2.78   55. male   no          no
6 1017.  0.483           2.78   96. male   no          no
# ... with 2 more variables: metastases <fct>, resection <chr>

tumor %>% group_by(sex) %>%
  make_newdata(age=seq(50,60,by=5), resection=unique(resection))
```

```
# A tibble: 12 x 9
    days status charlson_score   age sex    transfusion complications
   <dbl>  <dbl>          <dbl> <dbl> <fct>  <fct>       <fct>
 1 1060.  0.483           2.96   50. male   no          no
 2  954.  0.484           2.52   50. female no          no
 3 1060.  0.483           2.96   55. male   no          no
 4  954.  0.484           2.52   55. female no          no
 5 1060.  0.483           2.96   60. male   no          no
 6  954.  0.484           2.52   60. female no          no
 7 1060.  0.483           2.96   50. male   no          no
 8  954.  0.484           2.52   50. female no          no
 9 1060.  0.483           2.96   55. male   no          no
10  954.  0.484           2.52   55. female no          no
11 1060.  0.483           2.96   60. male   no          no
12  954.  0.484           2.52   60. female no          no
# ... with 2 more variables: metastases <fct>, resection <fct>
```

```
# same can be performed on ped data
make_newdata(ped_tumor, age=seq_range(age, n=3))
```

```
# A tibble: 3 x 14
  tstart  tend intlen interval    id offset ped_status charlson_score
   <dbl> <dbl>  <dbl> <fct>    <dbl>  <dbl>      <dbl>          <dbl>
1     0.    1.     1. (0,1]     393.     0.         0.           2.73
2     0.    1.     1. (0,1]     393.     0.         0.           2.73
3     0.    1.     1. (0,1]     393.     0.         0.           2.73
# ... with 6 more variables: age <dbl>, sex <fct>, transfusion <fct>,
#   complications <fct>, metastases <fct>, resection <fct>
```

```
# note that other interval related variables are adjusted as well
make_newdata(ped_tumor, tend=unique(tend)[1:4])
```

```
  tstart tend intlen interval       id    offset ped_status charlson_score
1      0    1      1    (0,1] 392.6801 0.0000000          0        2.72929
2      1    2      1    (1,2] 392.6801 0.0000000          0        2.72929
3      2    3      1    (2,3] 392.6801 0.0000000          0        2.72929
4      3    5      2    (3,5] 392.6801 0.6931472          0        2.72929
       age  sex transfusion complications metastases resection
1 61.31348 male          no            no        yes        no
2 61.31348 male          no            no        yes        no
3 61.31348 male          no            no        yes        no
4 61.31348 male          no            no        yes        no
```

```
ped_tumor %>% group_by(transfusion) %>% make_newdata(tend=unique(tend)[1:2])
```

```
  tstart tend intlen interval       id offset ped_status charlson_score
1      0    1      1    (0,1] 400.6291      0          0       2.684915
2      0    1      1    (0,1] 375.0737      0          0       2.827576
3      1    2      1    (1,2] 400.6291      0          0       2.684915
4      1    2      1    (1,2] 375.0737      0          0       2.827576
```

```
      age  sex transfusion complications metastases resection
1 61.3695 male          no            no        yes        no
2 61.1894 male         yes            no        yes        no
3 61.3695 male          no            no        yes        no
4 61.1894 male         yes            no        yes        no
```

Adding hazards, cumulative hazards and survival probabilities

Using these flexibly created new data sets, we employ **mgcv**'s `predict` function to calculate estimated log-hazards as well as secondary quantities like conditional survival probabilities from an estimated PAMM model (see also `?add_term`):

- hazard (`add_hazard`)/log-hazard (`add_hazard(..., type = "link")`):

```
new_df <- make_newdata(ped_tumor, tend = unique(tend)) %>% slice(1:5)
new_df %>% add_hazard(pam_tumor_tve, type = "link") %>%
  select(tend, hazard:ci_upper)

# A tibble: 5 x 5
   tend hazard    se ci_lower ci_upper
  <dbl>  <dbl> <dbl>    <dbl>    <dbl>
1    1.  -8.31 0.171    -8.65    -7.97
2    2.  -8.31 0.171    -8.65    -7.97
3    3.  -8.31 0.170    -8.65    -7.97
4    5.  -8.31 0.170    -8.65    -7.97
5    6.  -8.31 0.170    -8.65    -7.97
```

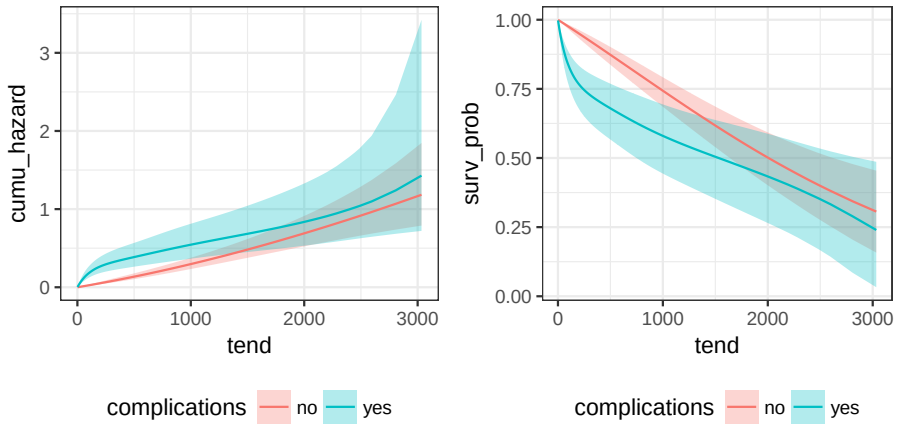
- cumulative hazard (`add_cumu_hazard`):

```
new_df %>% add_cumu_hazard(pam_tumor_tve) %>% add_surv_prob(pam_tumor_tve) %>%
  select(interval, cumu_hazard:surv_lower)

# A tibble: 5 x 7
  interval cumu_hazard cumu_lower cumu_upper surv_prob surv_upper
  <fct>          <dbl>      <dbl>      <dbl>     <dbl>      <dbl>
1 (0,1]       0.000246   0.000175   0.000346     1.000      1.000
2 (1,2]       0.000492   0.000350   0.000693     1.000      1.000
3 (2,3]       0.000739   0.000525   0.00104      0.999      0.999
4 (3,5]       0.00123    0.000876   0.00173      0.999      0.999
5 (5,6]       0.00148    0.00105    0.00208      0.999      0.999
# ... with 1 more variable: surv_lower <dbl>
```

Thus, the `add_*` functions add the calculated quantities directly to the data. The resulting augmented data sets can then be used for visualizations:

```
new_df <- ped_tumor %>%
  make_newdata(tend=unique(tend), complications=unique(complications)) %>%
  group_by(complications) %>%
  add_cumu_hazard(pam_tumor_tve) %>%
  add_surv_prob(pam_tumor_tve)
p_cumu <- ggplot(new_df, aes(x = tend, y = cumu_hazard, fill = complications,
    ymin = cumu_lower, ymax = cumu_upper)) +
  geom_ribbon(alpha = 0.3) + geom_line(aes(col = complications)) +
  theme(legend.position = "bottom")
p_surv <- p_cumu + aes(y = surv_prob, ymin = surv_lower, ymax = surv_upper)
gridExtra::grid.arrange(p_cumu, p_surv, nrow=1L)
```



5. Implementation details

In our implementation, we follow the principles of *tidy* data analysis (Wickham 2014), which implies that most functions take a data set as their first argument and all plot convenience functions are accompanied by respective functions that return the data used for plotting in a tidy format. All graphics in this article have been created using **ggplot2** (Wickham 2016b) and the visualization functions in **pammtols** also return `ggplot`-objects. Internally and in example code, we use **dplyr** (Wickham, Francois, Henry, and Müller 2017) and **tidyr** (Wickham 2016a) for data manipulation and **purrr** (Henry and Wickham 2018) for functional programming. **checkmate** (Lang 2017) and **testthat** (Wickham 2011) were used for defensive programming during the iterative development via **devtools** (Wickham, Hester, and Chang 2018). The flexible, formula based specification used to transform different data types to the PED format is facilitated by the **Formula** package (Zeileis and Croissant 2010). We compared the PAMM estimates to the Cox PH model, estimated using the `coxph` routine provided by the **survival** package (Therneau and Grambsch 2001), and to the Aalen model using the `aalen` routine provided by the **timereg** package (Martinussen and Scheike 2006). Simulation of time-to-event data from the $PEXP$ distribution is facilitated by the **msm** package (Jackson 2011). The companion website (https://adibender.github.io/pammtools/) was created using **pkgdown** (Wickham and Hesselberth 2018). This article was compiled using **knitr** (Xie 2015) based on **pammtools** v0.1.2 (Bender and Scheipl 2018).

6. Discussion

Summary

The **R** package **pammtools** facilitates the estimation, interpretation and visualization of flexible time-to-event regression analysis using GAMMs. In particular, in Section 3 we demonstrate how data of different complexity, including data with time-dependent covariates, can be transformed into a format suitable for such analyses. Special attention was given to the modeling and interpretation of time-varying effects (cf. Section 4.2) and cumulative effects (cf. Sections 3.2 and 4.3). In addition, Supplement A demonstrates how time-to-event data with complex time-varying and cumulative effects can be simulated, which will simplify future research on complex time-to-event models.

Limitations

Currently the package only supports data transformation for right-censored time-to-event data. While the PED format created by the `as_ped` function could be provided to any function or statistical software distribution that supports estimation of Poisson GA(M)Ms, most post-processing functions and convenience plot functions are customized to work with the **R** package **mgcv**. Although much effort went into making the respective functions robust, these efforts are limited by the fact that the estimation process is currently performed outside of **pammtools**. Feedback, bug reports and feature requests are welcomed at https://github.com/adibender/pammtools/issues or by contacting the authors.

Outlook

Future releases of **pammtools** will primarily focus on further improvement of the user interface and robustness of the implementation. We plan to extend the current framework to allow different censoring and truncation scenarios (left-truncation, left-censoring), as well as to support more complex outcomes like competing risk events or multi-state models.

References

Argyropoulos C, Unruh ML (2015). "Analysis of Time to Event Outcomes in Randomized Controlled Trials by Generalized Additive Models." *PLoS ONE*, **10**(4), e0123784. doi:10.1371/journal.pone.0123784. URL http://dx.doi.org/10.1371/journal.pone.0123784.

Bender A, Groll A, Scheipl F (2018a). "A generalized additive model approach to time-to-event analysis." *Statistical Modelling*, p. 1471082X17748083. ISSN 1471-082X. doi:10.1177/1471082X17748083. URL https://doi.org/10.1177/1471082X17748083.

Bender A, Scheipl F (2018). "adibender/pammtools: Piece-wise exponential Additive Mixed Modeling tools." URL https://doi.org/10.5281/zenodo.1258193.

Bender A, Scheipl F, Hartl W, Day AG, Küchenhoff H (2018b). "Penalized estimation of complex, non-linear exposure-lag-response associations." *Biostatistics*. doi:10.1093/biostatistics/kxy003. URL https://academic.oup.com/biostatistics/advance-article/doi/10.1093/biostatistics/kxy003/4852816.

Cai T, Hyndman RJ, Wand MP (2002). "Mixed Model-Based Hazard Estimation." *Journal of Computational and Graphical Statistics*, **11**(4), 784–798. ISSN 1061-8600. doi:10.1198/106186002862. URL http://dx.doi.org/10.1198/106186002862.

Demarqui FN, Loschi RH, Colosimo EA (2008). "Estimating the grid of time-points for the piecewise exponential model." *Lifetime Data Analysis*, **14**(3), 333–356. ISSN 1380-7870. doi:10.1007/s10985-008-9086-0.

Friedman M (1982). "Piecewise Exponential Models for Survival Data with Covariates." *The Annals of Statistics*, **10**(1), 101–113. ISSN 00905364. URL http://www.jstor.org/stable/2240502.

Gasparrini A (2014). "Modeling exposure-lag-response associations with distributed lag non-linear models." *Statistics in Medicine*, **33**(5), 881–899. ISSN 1097-0258. doi:10.1002/sim.5963. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.5963/abstract.

Henry L, Wickham H (2018). "**purrr**: Functional Programming Tools." URL https://CRAN.R-project.org/package=purrr.

Hofner B, Mayr A, Robinzonov N, Schmid M (2012). "Model-based boosting in **R**: a hands-on tutorial using the **R** package **mboost**." *Computational Statistics*, **29**(1-2), 3–35. ISSN 0943-4062, 1613-9658. doi:10.1007/s00180-012-0382-5. URL http://link.springer.com/article/10.1007/s00180-012-0382-5.

Holford TR (1980). "The Analysis of Rates and of Survivorship Using Log-Linear Models." *Biometrics*, **36**(2), 299–305. ISSN 0006341X. doi:10.2307/2529982. URL http://www.jstor.org/stable/2529982.

Hothorn T, Bühlmann P (2006). "Model-based boosting in high dimensions." *Bioinformatics*, **22**(22), 2828–2829. ISSN 1367-4803, 1460-2059. doi:10.1093/bioinformatics/btl462. URL http://bioinformatics.oxfordjournals.org/content/22/22/2828.

Jackson C (2011). "Multi-State Models for Panel Data: The **msm** Package for **R**." *http://www.jstatsoft.org/v38/i08/paper*.

Kauermann G (2005). "A note on smoothing parameter selection for penalized spline smoothing." *Journal of Statistical Planning and Inference*, **127**(1-2), 53–69. ISSN 0378-3758. doi:10.1016/j.jspi.2003.09.023. URL http://www.sciencedirect.com/science/article/pii/S0378375803003124.

Klein JP, Moeschberger ML (1997). *Survival analysis: Techniques for censored and truncated data*. Springer, New York. ISBN 978-0-387-94829-4.

Laird N, Olivier D (1981). "Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques." *Journal of the American Statistical Association*, **76**(374), 231–240. doi:10.2307/2287816. URL http://www.jstor.org/stable/2287816.

Lang M (2017). "**checkmate**: Fast Argument Checks for Defensive **R** Programming." *The R Journal*, **9**(1), 437–445. URL https://journal.r-project.org/archive/2017/RJ-2017-028/index.html.

Martinussen T, Scheike TH (2006). *Dynamic Regression Models for Survival Data*. Auflage: 2006 edition. Springer, New York, N.Y. ISBN 978-0-387-20274-7.

Sylvestre MP, Abrahamowicz M (2009). "Flexible modeling of the cumulative effects of time-dependent exposures on the hazard." *Statistics in Medicine*, **28**(27), 3437–3453. ISSN 1097-0258. doi:10.1002/sim.3701. URL http://onlinelibrary.wiley.com/doi/10.1002/sim.3701/abstract.

Therneau TM, Grambsch PM (2001). *Modeling Survival Data: Extending the Cox Model*. 1st ed. 2000. corr. 2nd printing 2001 edition. Springer, New York. ISBN 978-0-387-98784-2.

Whitehead J (1980). "Fitting Cox's Regression Model to Survival Data using GLIM." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **29**(3), 268–275. ISSN 0035-9254. doi:10.2307/2346901. URL http://www.jstor.org/stable/2346901.

Wickham H (2011). "**testthat**: Get Started with Testing." *The R Journal*, **3**, 5–10. URL http://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Wickham H (2014). "Tidy Data." *Journal of Statistical Software*, **59**(10). ISSN 1548-7660. doi:10.18637/jss.v059.i10. URL http://www.jstatsoft.org/v59/i10/.

Wickham H (2016a). *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016 edition. Springer, New York, NY. ISBN 978-3-319-24275-0.

Wickham H (2016b). ***tidyr***: *Easily Tidy Data with* `spread()` *and* `gather()` *Functions*. URL https://CRAN.R-project.org/package=tidyr.

Wickham H, Francois R, Henry L, Müller K (2017). "**dplyr**: A Grammar of Data Manipulation." URL https://CRAN.R-project.org/package=dplyr.

Wickham H, Hesselberth J (2018). "**pkgdown**: Make Static HTML Documentation for a Package." URL http://pkgdown.r-lib.org.

Wickham H, Hester J, Chang W (2018). "**devtools**: Tools to Make Developing **R** Packages Easier." URL https://CRAN.R-project.org/package=devtools.

Wood S, Scheipl F (2017). "**gamm4**: Generalized Additive Mixed Models using **mgcv** and **lme4**." URL https://CRAN.R-project.org/package=gamm4.

Wood SN (2011). "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(1), 3–36. ISSN 13697412. doi:10.1111/j.1467-9868.2010.00749.x.

Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2 rev ed. edition. Chapman & Hall/Crc Texts in Statistical Science, Boca Raton. ISBN 978-1-4987-2833-1.

Wood SN, Goude Y, Shaw S (2015). "Generalized additive models for large data sets." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **64**(1), 139–155. ISSN 1467-9876. doi:10.1111/rssc.12068. URL http://onlinelibrary.wiley.com/doi/10.1111/rssc.12068/abstract.

Xie Y (2015). *Dynamic documents with **R** and **knitr***. Second edition edition. CRC Press/Taylor & Francis, Boca Raton. ISBN 978-1-4987-1696-3.

Zeileis A, Croissant Y (2010). "Extended Model Formulas in **R** : Multiple Parts and Multiple Responses." *Journal of Statistical Software*, **34**(1). ISSN 1548-7660. doi:10.18637/jss.v034.i01. URL http://www.jstatsoft.org/v34/i01/.

A. Simulating time-to-event data

For convenience, the **pammtools** package contains a lightweight, but versatile function for the simulation of time-to-event data, with potentially smooth, smoothly time-varying effects. For the simulation of survival times we use the Piece-wise exponential distribution $t \sim \text{PEXP}(\boldsymbol{\lambda}_i, \mathbf{t})$, which is implemented in the **R** package **msm** (Jackson 2011) Here $\boldsymbol{\lambda}$ is a vector of hazards at time points $\mathbf{t}$ and $\boldsymbol{\lambda}$ can be specified conveniently using a `formula` notation.

In Section A.1, we empirically demonstrate that even crude PEXP hazards can be used to simulate survival times from continuous distributions. In Section A.2 we illustrate the simulation of survival times based on hazard rates that flexibly depend on time-constant covariates. Lastly, Section A.3 shows how to simulate from hazards with cumulative effects of TDCs.

A.1. Motivation

We use a simple Weibull baseline hazard model to illustrate that the function indeed simulates event times from the desired distribution, even though the hazards $\boldsymbol{\lambda}$ are assumed to be piece-wise constant between two time-points in $\mathbf{t}$. Figure 10 depicts the hazard rate and survivor function of a Weibull distribution with $T \sim WB(\alpha = 1.5, \lambda = 10)$.

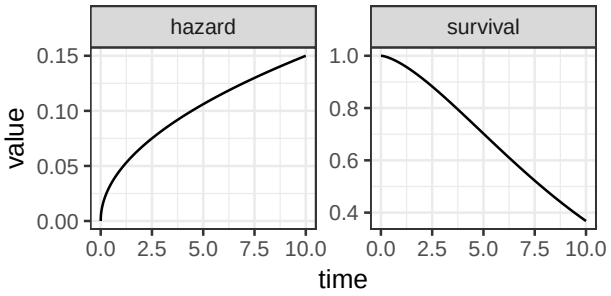


Figure 10: Hazard rate (left) and survivor function (right) of the $WB(1.5, 10)$ distribution.

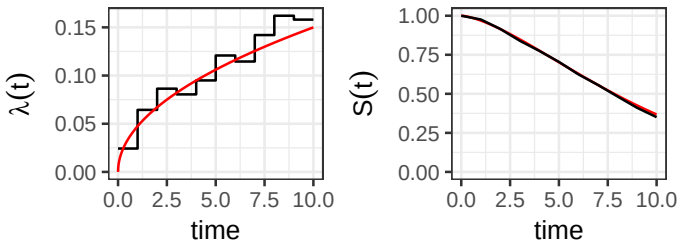
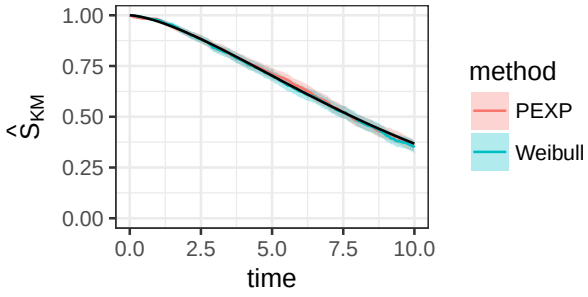


Figure 11: PEM estimates of the baseline hazard $\lambda(t)$ (left panel) and survival probability $S(t)$ (right panel). Red lines indicate the true Weibull hazard and survival probability, respectively.

Figure 11 (left panel) shows the baseline hazard estimated by a PEM with 10 intervals based

on $n = 1000$ survival times simulated from $WB(1.5, 10)$. Although the approximation of the underlying smooth hazard is relatively crude, the survival function calculated from this step hazard is very close to the true survivor function (cf. right panel of Figure 11). Finally, Figure 12 depicts the distribution of survival times (Kaplan-Meier estimates) for $n = 1000$ survival times simulated directly from the correct Weibull distribution (`rweibull(n, 1.5, 10)`) on the one hand and from the $PEXP$ distribution (based on the crude hazard in Figure 11) on the other hand.

Figure 12: Comparison of Kaplan-Meier survival probability estimates based on survival times simulated directly from the Weibull distribution $WB(1.5, 10)$ and based on survival times simulated from the $PEXP$ distribution based on the hazards depicted in Figure 11. The Black line indicates the true Weibull survival probability on $t \in [0, 10]$.



A.2. Flexible, covariate dependent simulation of survival times

To simulate survival times from the PEXP distribution conveniently, **pammtools** provides the `sim_pexp` function. Similar to the `as_ped` function, it uses a formula interface, which allows to specify complex hazards relatively easily. For example, in **R**-chunk 16 we simulate data from

$$\log(\lambda(t|x_1, x_2)) = -3.5 + f_0(t) - 0.5x_1 + \sqrt{x_2},$$

where $f_0(t)$ is a Gamma(8,2) density function. Any existing or previously defined function can be used in the `formula` argument to `sim_pexp`. The argument `cut` defines the time-points at which the piece-wise constant hazard will change its value. In R chunk 16 for example, the hazard will change its value at $t = 1, t = 2, \ldots$ with $f_0(t)$ (and other time-varying effects) evaluated at the respective interval end-points. `sim_pexp` returns the original data augmented by the simulated survival times (`time`) as well as a `status` column.

R-chunk 16

```
# basic data
set.seed(7042018)
# create data set with covariates
n <-1000
df <- tibble::tibble(x1 = runif(n, -3, 3), x2 = runif(n, 0, 6))
# baseline hazard function
f0 <- function(t) {dgamma(t, 8, 2) * 6}
# simulate data from PEXP
```

```
sim_df <- sim_pexp(
  formula = ~ -3.5 + f0(t) -0.5*x1 + sqrt(x2),
  data    = df,
  cut     = 0:10)
```

Note that the simulation could be easily extended to contain time-varying effects, e.g. by defining a function

```
f_tx <- function(t, x) sqrt(x)*log(t)
```

and calling

```
sim_pexp(~ -3.5 + f0(t) -0.5*x1 + f_tx(t, x2), data = df, cut = 0:10)
```

A.3. Simulation of survival times with cumulative effects

Weighted cumulative exposure

In this section we demonstrated how to simulate data with hazard rate

$$\log(\lambda(t|x_1, x_2, \mathbf{z})) = -3.5 + f_0(t) - 0.5x_1 + \sqrt{x2} + \int_{\mathcal{T}(t)} h(t - t_z)z(t_z)\mathrm{d}t_z.$$

which constitutes a so-called Weighted cumulative exposure model (Sylvestre and Abrahamowicz 2009). This data is used in section 4.3.1 to illustrate estimation and visualizations of such effects. The static part of the data set as well as the baseline hazard and TCC effects are identical to the previous section (cf. **R**-chunk 16). For the cumulative effect, we define the exposure time grid (i.e., the time points t_z at which the TDC was observed) and use the function `add_tdc` (mnemonic: *add time-dependent covariate*) to add the information on the exposure times and the $z(t_z)$ to the data (cf. **R**-chunk 17).

R-chunk 17

```
# define follow-up time grid for simulation
# (arbitrary, but check that enough events are observed over follow-up)
time_grid <- seq(0, 10, by = 0.5)
# baseline hazard
f0 <- function(t) {dgamma(t, 8, 2) * 6}

# define time grid on which TDC is observed
# (arbitrary, but lag-lead matrix will depend on it)
tz <- seq(-5, 5, by = .25)
# define function that generates nz exposures z(t_{z,1}), ..., z(t_{z,Q})
rng_z = function(nz) {
  as.numeric(arima.sim(n = nz, list(ar = c(.8, -.1))))
}
## add TDCs to data set
df <- df %>% add_tdc(tz, rng_z)
```

```
df %>% slice(1) %>% pull("tz")

[[1]]
 [1] -5.00 -4.75 -4.50 -4.25 -4.00 -3.75 -3.50 -3.25 -3.00 -2.75 -2.50
[12] -2.25 -2.00 -1.75 -1.50 -1.25 -1.00 -0.75 -0.50 -0.25  0.00  0.25
[23]  0.50  0.75  1.00  1.25  1.50  1.75  2.00  2.25  2.50  2.75  3.00
[34]  3.25  3.50  3.75  4.00  4.25  4.50  4.75  5.00

# df %>% slice(1) %>% pull("z.tz")
```

The partial effect $h(t-t_z)z(t_z)$ (see function `f_wce`) and the lag-lead window $\mathcal{T}(t)$ (see function `ll_fun`) are defined in **R**-chunk 18 and depicted in Figure 13. The left panel of Figure 13 shows the latency-dependent weight function $h(t-t_z)$ for the exposures $z(t_z)$. The middle panel shows the lag-lead window with partial effects. Note that $h(t-t_z)$ only depends on the latency, not the specific combination of t and t_z. Nonetheless, the cumulative effect $g(\mathbf{z}, t)$ (right panel) varies over t even for constant exposure $z(t_z) = z$ since it is integrated over different windows of effectiveness $\mathcal{T}(t)$.

R-chunk 18

```
# define lag-lead function: integrate over the preceding 12 time units
ll_fun <- function(t, tz) ((t - tz) >= 0) & ((t - tz) <=  12)
# gg_laglead(0:10, -5:5, ll_fun)

# partial effect h(t - tz) * z
f_wce <- function(t, tz, z) {
  0.5 * (dnorm(t - tz, 6, 2.5)) * z
}
```

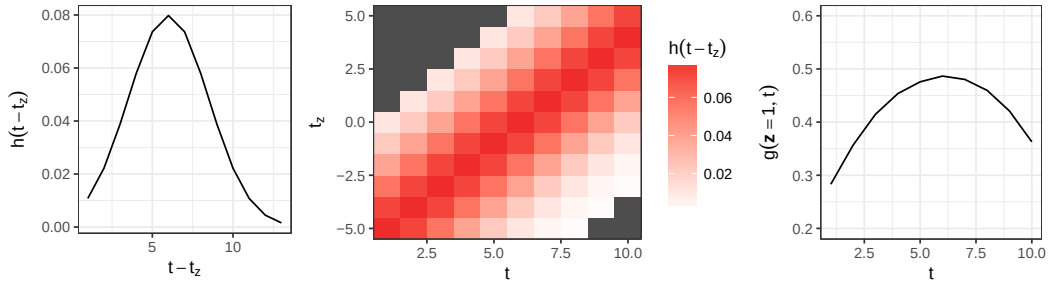


Figure 13: Left: Partial effect $h(t-t_z)$ defined in **R**-chunk 18 for different latencies $t-t_z$. Middle: The lag-lead window $\mathcal{T}(t)$ and respective partial effects for each combination of t and t_z. Combinations of t and t_z outside the specified lag-lead window in dark gray. Partial effects of exposures at different time-points t, t_z are the same if the latency $t-t_z$ is the same, i.e. $h(5-1) = h(6-2) = h(4)$. Right: Cumulative effect $g(\mathbf{z}, t)$ for constant $z(t_z) = 1 \,\forall\, t_z$.

Given the above setup with cumulative effects $g(\mathbf{z}, t) = \int_{\mathcal{T}(t)} h(t-t_z)z(t_z)\mathrm{d}t_z$, we can now simulate the data using the `sim_pexp` function as displayed in **R**-chunk 19.

R-chunk 19

```
simdf_wce <- sim_pexp(
  formula = ~ -3.5 + f0(t) -0.5*x1 + sqrt(x2)|
     fcumu(t, tz, z.tz, f_xyz=f_wce, ll_fun=ll_fun),
  data = df, cut = time_grid)
```

Bivariate, smooth partial effects

In this section we illustrate an extension of the previous simulation, where the exposure $z(t_z)$ affects the hazard non-linearly as denoted in eq. 9.

$$\log(\lambda(t|x_1, x_2, \mathbf{z})) = -3.5 + f_0(t) - 0.5x_1 + \sqrt{x2} + \int_{\mathcal{T}(t)} h(t - t_z, z(t_z))\mathrm{d}t_z \tag{9}$$

Using the `sim_pexp` function, we can extend the previous simulation (cf. Section A.3.1) by changing the partial effect function as illustrated in **R**-chunk 20 (function `f_dlnm`). Figure 14 depicts the bivariate, smooth partial effect $h(t-t_z, z(t_z))$ and the resulting cumulative effects $g(\mathbf{z}, t)$ for a simplified exposure history with constant $z(t_z) = 1$ all t_z.

R-chunk 20

```
# partial effect h(t - tz) * z
f_dlnm <- function(t, tz, z) {
   20 * ((dnorm(t - tz, 6, 2.5)) * (dnorm(z, 1.25, 2.5) - dnorm(-1, 1.25, 2.5)))
}

simdf_dlnm <- sim_pexp(
  formula = ~ -4.5 + f0(t) -0.5*x1 + sqrt(x2)|
     fcumu(t, tz, z.tz, f_xyz=f_dlnm, ll_fun=ll_fun),
  data = df, cut = time_grid)
```

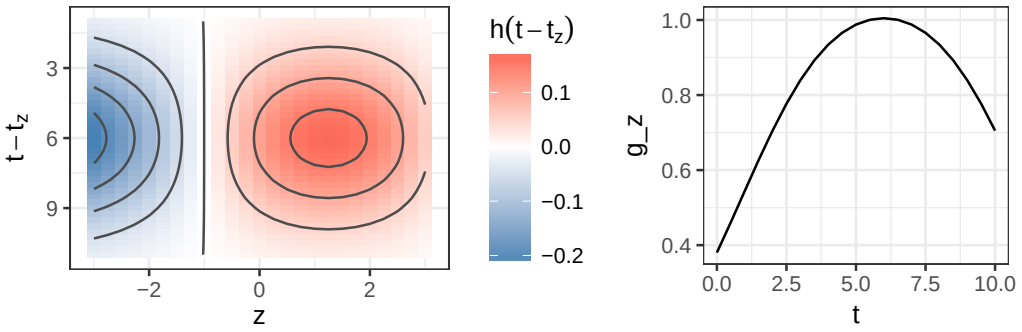


Figure 14: **Left**: Partial effect $h(t - t_z, z(t_z))$ used for the simulation of survival times (data `simdf_dlnm`) in **R**-chunk 20. **Right**: The cumulative effects $g(\mathbf{z}, t)$ resulting from constant exposure histories $z(t_z) = 1 \,\forall\, t_z$.

Bivariate smooth of time and exposure time

Here we simulate the data used in Section 4.3 with hazard

$$\log(\lambda(t|x_1, x_2, \mathbf{z})) = -3.5 + f_0(t) - 0.5x_1 + \sqrt{x2} + \int_{\mathcal{T}(t)} h(t, t_z)z(t_z)\mathrm{d}t_z.$$

The simulation code is given in **R**-chunk 21 with updated partial effect function `f_elra`. Figure 15 depicts the bivariate, smooth partial effect $h(t.t_z)$ (left panel) and the resulting cumulative effect $g(\mathbf{z}, t)$ for a simplified exposure history with $z(t_z) = 1 \forall t_z$ (right panel).

R-chunk 21

```
# partial effect h(t,tz) * z
f_elra <- function(t, tz, z) {
   5*(-(dnorm(tz, -1, 2.5)) * (dnorm(t, 5, 1.5) - dnorm(5, 5, 1.5)))*z
}
simdf_tv_wce <- sim_pexp(formula = ~ -4.5 + f0(t) -0.5*x1 + sqrt(x2)|
     fcumu(t, tz, z.tz, f_xyz = f_elra, ll_fun = ll_fun),
  data = df, cut = time_grid)
```

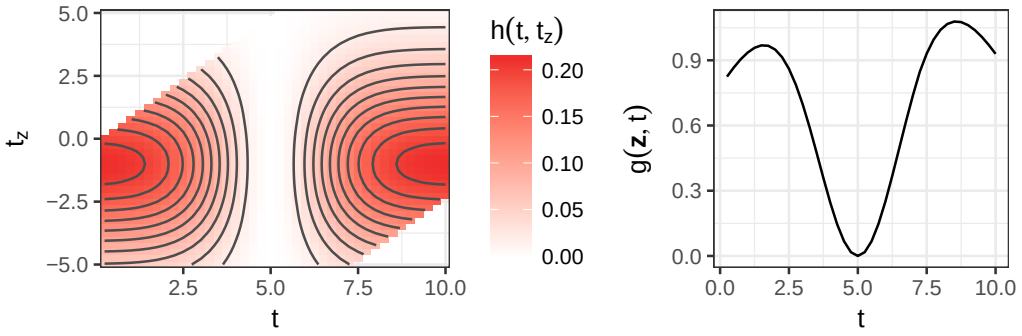


Figure 15: Left: Bivariate partial effect surface $h(t, t_z)$, combinations of t and t_z that lie outside the lag-lead window $\mathcal{T}(t)$ are omitted. Right: The cumulative effect resulting from the partial effect depicted in the left panel for a simplified exposure profile with $z(t_z) = 1 \forall t_z$.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12. Juli 2011, §8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 05.06.2018 Andreas Bender